

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:50:29 ; Search time 31 Seconds
(without alignments)
94.565 Million cell updates/sec

Title: US-09-929-788-3
Perfect score: 121
Sequence: 1 FFHHIFRGIHVGVKTIHKLVTG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 6
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDs2/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDs2/gcgdata/geneeqp-emb1/AA1981.DAT:*
3: /SIDs2/gcgdata/geneeqp-emb1/AA1982.DAT:*
4: /SIDs2/gcgdata/geneeqp-emb1/AA1983.DAT:*
5: /SIDs2/gcgdata/geneeqp-emb1/AA1984.DAT:*
6: /SIDs2/gcgdata/geneeqp-emb1/AA1985.DAT:*
7: /SIDs2/gcgdata/geneeqp-emb1/AA1986.DAT:*
8: /SIDs2/gcgdata/geneeqp-emb1/AA1987.DAT:*
9: /SIDs2/gcgdata/geneeqp-emb1/AA1988.DAT:*
10: /SIDs2/gcgdata/geneeqp-emb1/AA1989.DAT:*
11: /SIDs2/gcgdata/geneeqp-emb1/AA1990.DAT:*
12: /SIDs2/gcgdata/geneeqp-emb1/AA1991.DAT:*
13: /SIDs2/gcgdata/geneeqp-emb1/AA1992.DAT:*
14: /SIDs2/gcgdata/geneeqp-emb1/AA1993.DAT:*
15: /SIDs2/gcgdata/geneeqp-emb1/AA1994.DAT:*
16: /SIDs2/gcgdata/geneeqp-emb1/AA1995.DAT:*
17: /SIDs2/gcgdata/geneeqp-emb1/AA1996.DAT:*
18: /SIDs2/gcgdata/geneeqp-emb1/AA1997.DAT:*
19: /SIDs2/gcgdata/geneeqp-emb1/AA1998.DAT:*
20: /SIDs2/gcgdata/geneeqp-emb1/AA1999.DAT:*
21: /SIDs2/gcgdata/geneeqp-emb1/AA2000.DAT:*
22: /SIDs2/gcgdata/geneeqp-emb1/AA2001.DAT:*
23: /SIDs2/gcgdata/geneeqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	22	23	AAE20174
2	121	100.0	22	23	ABB07722
3	118	97.5	22	23	AAE20173
4	118	97.5	22	23	ABB07721
5	84	69.4	22	23	AAE20172
6	84	69.4	22	23	ABB07720
7	65	53.7	24	23	AAE20177
8	51	42.1	372	14	AAE47051
9	51	42.1	436	16	AAE74803
10	51	42.1	436	17	AAE92519

11	51	42.1	436	18	AAE23219	Protein product of
12	51	42.1	436	22	AAE90933	C glutamicum prote
13	51	42.1	436	22	AAE92342	C glutamicum prote
14	51	42.1	436	22	AAE92467	C glutamicum prote
15	48	39.7	78	22	AAE60341	Protonibacterium
16	47	38.8	435	18	AAE23218	Protein product of
17	47	38.8	612	23	AAE47617	Listeria monocytog
18	46	38.0	44	23	AAE20175	Endobiotic family
19	46	38.0	44	23	AAE07723	Antimicrobial pept
20	46	38.0	384	22	AAE92653	C glutamicum prote
21	46	38.0	388	22	AAE79449	Corynebacterium gl
22	46	38.0	388	22	AAE79450	Corynebacterium gl
23	46	38.0	907	22	AAE6197	Putative P. abyssi
24	45.5	37.6	290	21	AAE40499	Human ORFX ORF23
25	45.5	37.6	392	22	AAE85392	Stem cell growth f
26	45.5	37.6	392	22	AAE85393	Stem cell growth f
27	45.5	37.6	425	22	AAE64527	Human polypeptide
28	45.5	37.6	425	22	AAE85399	Stem cell growth f
29	45.5	37.6	427	22	AAE40853	Human polypeptide
30	45.5	37.6	427	22	AAE40854	Human polypeptide
31	45.5	37.6	449	22	AAE85398	Stem cell growth f
32	45.5	37.6	486	22	AAE39067	Human polypeptide
33	45.5	37.6	499	22	AAE85396	Stem cell growth f
34	45.5	37.6	529	22	AAE38068	Human polypeptide
35	45.5	37.6	529	22	AAE85394	Stem cell growth f
36	45.5	37.6	529	22	AAE85394	Amino acid sequenc
37	45.5	37.6	529	22	AAE31211	Human Tumour Endot
38	45.5	37.6	529	23	AAE90726	Human Tumour Endot
39	45.5	37.6	529	23	AAE90734	Human immune/haema
40	45	37.2	32	22	AAE82596	Human ORFX protein
41	45	37.2	92	23	ABP03651	Novel human diagno
42	45	37.2	145	22	ABG16238	Human protein sequ
43	45	37.2	269	22	AAE95041	C glutamicum prote
44	45	37.2	380	22	AAE92198	C glutamicum prote
45	45	37.2	380	22	AAE79495	Corynebacterium gl

ALIGNMENTS

RESULT 1	AAE20174	standard; peptide; 22 AA.
ID	AAE20174	
XX	AAE20174	
AC	AAE20174	
XX	18-JUN-2002	(first entry)
DE	Fish mast cell antimicrobial peptide, piscidin 2 (P2).	
XX		
KW	Antimicrobial peptide; endobiotic peptide; mast cell; piscidin; fish;	
KW	methicillin; vancomycin; streptogramin; microbial infection; stress;	
KW	bacterial disease; fish health; screening; gene therapy; antimicrobial;	
KW	tranquilliser; fungicide; protozoacide; antibacterial; antiparasitic.	
OS	Morone chryseops.	
XX	Morone saxatilis.	
PN	WO200214345-A2.	
XX		
PD	21-FEB-2002.	
XX		
PF	13-AUG-2001; 2001WO-US41696.	
XX		
PR	15-AUG-2000; 2000US-225354P.	
PA	(UNCL-) UNIV NORTH CAROLINA STATE.	
XX		
PI	Noga EJ, Silphaduang U;	
XX		
DR	WPI; 2002-269176/31.	
XX		
PT	Novel antimicrobial peptides (endobiotic peptides) isolated from mast	

CC product. Nucleic acid sequences encoding these peptides are useful for
CC diagnostic purposes, treating stress in a fish and for screening
CC additional endobiotic peptides. They are also useful in gene therapy.
CC The present sequence is piscidin 1 (P1) peptide isolated from the gills
CC of hybrid striped bass (Morone saxatilis x Morone chrysops).

XX Sequence 22 AA;

Query Match 97.5%; Score 118; DB 23; Length 22;
Best Local Similarity 95.5%; Pred. No. 1.6e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FFHHIFRGIVHGKTIHKLVTG 22
Db 1 FFHHIFRGIVHGKTIHKLVTG 22

RESULT 4

ABB07721 ABB07721 standard; peptide; 22 AA.

XX ABB07721;

XX 10-JUN-2002 (first entry)

XX Antimicrobial peptide #2 from hybrid striped bass.

XX Fish; hybrid striped bass; endobiotic; antibacterial; tranquilizer;
XX antimicrobial; antibacterial; fish food product.

OS Morone saxatilis x Morone chrysops.

XX WO200214346-A2.

XX 21-FEB-2002.

XX 13-AUG-2001; 2001WO-US41697.

XX 15-AUG-2000; 2000US-225354P.

XX (UYNC-) UNIV NORTH CAROLINA STATE.

XX Noga EJ, Silphaduang U;

XX WPI; 2002-269177/31.

XX Novel antimicrobial compound (endobiotic peptide isolated from fish),
XX useful for monitoring fish health and monitoring freshness of fish food
XX product and for screening compounds which treat stress in fish -

XX Claim 1; Page 16; 25pp; English.

XX The invention provides antimicrobial peptides (ABB07720-23) isolated from
XX hybrid striped bass. The peptides were purified from the endobiotic
XX family. The endogenous endobiotic peptides exhibiting antimicrobial
XX activity are useful for monitoring fish health, monitoring the freshness
XX of the fish product such as fresh, refrigerated, or frozen, fish food
XX product, and for screening compounds useful for treating stress in fish.
XX Nucleic acids encoding the antimicrobial peptides are useful for treating
XX stress in a fish. Sequences ABB07720-23 represent specific examples of
XX the antimicrobial peptides.

XX Sequence 22 AA;

Query Match 97.5%; Score 118; DB 23; Length 22;
Best Local Similarity 95.5%; Pred. No. 1.6e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FFHHIFRGIVHGKTIHKLVTG 22
Db 1 FFHHIFRGIVHGKTIHKLVTG 22

RESULT 5
AAE20172 AAE20172 standard; peptide; 22 AA.

XX AAE20172;

XX 18-JUN-2002 (first entry)

XX Fish mast cell antimicrobial peptide, piscidin 3 (P3).

XX Antimicrobial peptide; endobiotic peptide; mast cell; piscidin; fish;
XX methicillin; vancomycin; streptogramin; microbial infection; stress;
XX bacterial disease; fish health; screening; gene therapy; antimicrobial;
XX tranquilizer; fungicide; protozoicide; antibacterial; antiparasitic.

OS Morone chrysops.

XX Morone saxatilis.

XX WO200214345-A2.

XX 21-FEB-2002.

XX 13-AUG-2001; 2001WO-US41696.

XX 15-AUG-2000; 2000US-225354P.

XX (UYNC-) UNIV NORTH CAROLINA STATE.

XX Noga EJ, Silphaduang U;

XX WPI; 2002-269176/31.

XX Novel antimicrobial peptides (endobiotic peptides) isolated from mast
XX cells, specifically fish mast cells, useful for treating microbial
XX infections in humans or animals and for reducing antibiotic resistance
XX in bacteria -

XX Claim 4; Page 17; 31pp; English.

XX The invention relates to antimicrobial peptide (endobiotic peptide)
XX isolated from a mast cell. Such peptides obtained from fish mast cells
XX are referred to as piscidins. The antimicrobial peptide is useful for
XX treating microbial infection in a subject, and for reducing antibiotic
XX (e.g., methicillin, vancomycin or streptogramin) resistance in bacteria
XX such as *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus aureus*,
XX *Klebsiella pneumoniae*, *Streptococcus faecalis* or *Shigella flexneri*. These
XX peptides are useful for treating microbial infections in human or animals
XX and for preparing a medicament for treating microbial infections. These
XX peptides are also useful for treating infections caused by microorganisms
XX other than bacteria, e.g., fungi, yeast, protozoa, parasite, etc. These
XX peptides are useful for treating stress or bacterial disease in fish and
XX monitoring fish health. The inverse relationship between endobiotic
XX levels and stress also allows assessment of freshness of fish food
XX product. Nucleic acid sequences encoding these peptides are useful for
XX diagnostic purposes, treating stress in a fish and for screening
XX additional endobiotic peptides. They are also useful in gene therapy.
XX The present sequence is piscidin 3 (P3) peptide isolated from the gills
XX of hybrid striped bass (Morone saxatilis x Morone chrysops).

XX Sequence 22 AA;

Query Match 69.4%; Score 84; DB 23; Length 22;
Best Local Similarity 63.6%; Pred. No. 3.1e-06;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 FFHHIFRGIVHGKTIHKLVTG 22
Db 1 FTHHIFRGIVHAGRSIGRFLTG 22

RESULT 6

ABB07720 ABB07720 standard; peptide; 22 AA.

```

XX ABB07720;
AC 10-JUN-2002 (first entry)
DE Antimicrobial peptide #1 from hybrid striped bass.
XX
XX Fish; hybrid striped bass; endobiotic; antibiotic; tranquilizer;
XX antimicrobial; antibacterial; fish food product.
XX
XX Morone saxatilis x Morone chrysops.
XX
XX WO200214346-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US41697.
XX
XX 15-AUG-2000; 2000US-225354P.
XX
XX (UNNC-) UNIV NORTH CAROLINA STATE.
XX
XX Noga EJ, Silphaduang U;
XX WPI; 2002-269177/31.
XX
XX Novel antimicrobial compound (endobiotic peptide isolated from fish).
XX useful for monitoring fish health and monitoring freshness of fish food
XX product and for screening compounds which treat stress in fish -
XX
XX Claim 1; Page 16; 25pp; English.
XX
XX The invention provides antimicrobial peptides (ABB07720-23) isolated from
XX hybrid striped bass. The peptides were purified from the endobiotic
XX family. The endogenous endobiotic peptides exhibiting antimicrobial
XX activity) are useful for monitoring fish health, monitoring the freshness
XX of the fish product such as fresh, refrigerated, or frozen, fish food
XX product, and for screening compounds useful for treating stress in fish.
XX Nucleic acids encoding the antimicrobial peptides are useful for treating
XX stress in a fish. Sequences ABB07720-23 represent specific examples of
XX the antimicrobial peptides.
XX
XX Sequence 22 AA;
SQ
Query Match 69.4%; Score 84; DB 23; Length 22;
Best Local Similarity 63.6%; Pred. No. 3.1e-06;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 FFHHIFRGIVHVKTIHLVLTG 22
DB 1 FHHHIFRGIVHAGRSIGRFLTG 22

```

RESULT 7
AAE20177
ID AAE20177 standard; peptide; 24 AA.
XX
AC AAE20177;
XX
DT 18-JUN-2002 (first entry)
XX
DE Antimicrobial peptide analogue.
XX
KM Antimicrobial peptide; endobiotic peptide; mast cell; piscidin;
KM methicillin; vancomycin; streptogramin; microbial infection; stress;
KM bacterial disease; fish health; screening; gene therapy; antimicrobial;
KM tranquilizer; fungicide; protozoicide; antibacterial; antiparasitic.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 12..13
XX

```

PN WO200214345-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US41696.
XX
XX 15-AUG-2000; 2000US-225354P.
XX
XX (UNNC-) UNIV NORTH CAROLINA STATE.
XX
XX Noga EJ, Silphaduang U;
XX WPI; 2002-269176/31.
XX
XX Novel antimicrobial peptides (endobiotic peptides) isolated from mast
XX cells, specifically fish mast cells, useful for treating microbial
XX infections in humans or animals and for reducing antibiotic resistance
XX in bacteria -
XX
XX Example 9; Page 24; 31pp; English.
XX
XX The invention relates to antimicrobial peptide (endobiotic peptide)
XX isolated from a mast cell. Such peptides obtained from fish mast cells
XX are referred as piscidins. The antimicrobial peptide is useful for
XX treating microbial infection in a subject, and for reducing antibiotic
XX (e.g., methicillin, vancomycin or streptogramin) resistance in bacteria
XX such as Pseudomonas aeruginosa, Escherichia coli, Staphylococcus aureus,
XX Klebsiella pneumoniae, Streptococcus faecalis or Shigella flexneri. These
XX peptides are useful for treating microbial infections in human or animals
XX and for preparing a medicament for treating microbial infections. These
XX peptides are also useful for treating infections caused by microorganisms
XX other than bacteria, e.g., fungi, yeast, protozoa, parasite, etc. These
XX peptides are useful for treating stress or bacterial disease in fish and
XX monitoring fish health. The inverse relationship between endobiotic
XX levels and stress also allows assessment of freshness of fish food
XX product. Nucleic acid sequences encoding these peptides are useful for
XX diagnostic purposes, treating stress in a fish and for screening
XX additional endobiotic peptides. They are also useful in gene therapy.
XX The present sequence is antimicrobial peptide analogue.
XX
XX Sequence 24 AA;
SQ
Query Match 53.7%; Score 65; DB 23; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FFHHIFRGIVH 11
DB 14 FHHHIFRGIVH 24

```

RESULT 8
AAR47051
ID AAR47051 standard; Protein; 372 AA.
XX
AC AAR47051;
XX
DT 19-APR-1994 (first entry)
XX
DE ISG1 ORF-2 prod.
XX
KM Insertion element; ISC; transposon; vector;
KM inverse repetitive end; pW5; sac-B gene.
XX
OS Corynebacterium glutamicum mutant LT 5.5.
XX
PN DE4208785-A.
XX
PD 23-SEP-1993.
XX
PF 19-MAR-1992; 92DE-4208785.
XX
PR 19-MAR-1992; 92DE-4208785.
XX

XX (DEGS) DEGUSA AG.
 PA Jueger W, Kalinowski J, Puehler A, Schaefer A, Seep-Feldhaus A;
 PI Wohlleben W;
 XX WPI; 1993-304217/39.
 DR N-PSDB; AAQ48811.
 XX
 PT Rapid detection of new insertion elements and transposons in
 PT coryneform bacteria - using new recombinable vector contg. sac-B
 PT gene, used e.g. for mutagenesis
 XX
 PS Disclosure; Fig 2; 15pp; German.
 CC
 CC pW5 contg. the sac-B gene was transferred from E.coli S17-1
 CC to recipient strains. 16 colonies of C. glutamicum had
 CC a 1.45 kb insert in the sac-B gene of pW5. All these inserts
 CC hybridised with a digoxigenin-dUTP labelled ISCG1 DNA
 CC (isolated from C. glutamicum mutant LTS-5).
 CC
 XX
 SQ Sequence 372 AA;
 Query Match 42.1%; Score 51; DB 14; Length 372;
 Best Local Similarity 41.2%; Pred. No. 8.5;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 3 HHIFRGIVHGKTIHKL 19
 DB 353 HYILRLIHSGQLVHKL 369
 RESULT 9
 AAR74803
 ID AAR74803 standard; Protein; 436 AA.
 XX
 AC AAR74803;
 XX
 DT 10-JAN-1996 (first entry)
 XX
 DE Corynebacterium glutamicum insertion sequence.
 XX
 XX Corynebacterium glutamicum; insertion sequence; gene mapping;
 KW promoter detection; insertion mutant preparation;
 KW insertion of gene information; interruption of genes.
 OS Corynebacterium glutamicum.
 PN JP07107976-A.
 XX
 XX 25-APR-1995.
 PD
 XX 15-OCT-1993; 93JP-0258459.
 PF
 XX 15-OCT-1993; 93JP-0258459.
 PR
 XX (MITP) MITSUBISHI PETROCHEMICAL CO LTD.
 PA
 XX WPI; 1995-190176/25.
 DR N-PSDB; AAQ90280.
 XX
 PT A new coryneform bacterium insertion sequence - can be used for the
 PT preparation of an insertion mutant
 XX
 PS Claim 1; Pages 7-9; 10pp; Japanese.
 XX
 CC AAQ90280 a Corynebacterium glutamicum insertion sequence encodes,
 CC AAR74803. The sequence can be used for the prepn. of an insertion
 CC mutant, gene mapping, promoter detection, insertion of gene
 CC information and the interruption of specified genes.
 CC
 XX Sequence 436 AA;
 SQ

Query Match 42.1%; Score 51; DB 16; Length 436;
 Best Local Similarity 41.2%; Pred. No. 10;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 3 HHIFRGIVHGKTIHKL 19
 DB 417 HYILRLIHSGQLVHKL 433
 RESULT 10
 AAR92519
 ID AAR92519 standard; Protein; 436 AA.
 XX
 AC AAR92519;
 XX
 DT 01-AUG-1996 (first entry)
 XX
 DE Corynebacterium glutamicum open reading frame protein.
 XX
 XX autonomous replication; inverted repeat; insertion sequence;
 KW open reading frame; plasmid; isolation; amino acid biosynthesis;
 KW marker gene.
 XX
 OS Coryneform bacterium.
 XX
 PN JP07327680-A.
 XX
 PD 19-DEC-1995.
 XX
 PF 07-JUN-1994; 94JP-0124852.
 XX
 XX 07-JUN-1994; 94JP-0124852.
 PR
 XX 07-JUN-1994; 94JP-0124852.
 XX
 PA (MITU) MITSUBISHI CHEM CORP.
 XX
 DR WPI; 1996-072338/08.
 DR N-PSDB; AAT16268-69.
 XX
 PT Plasmid which cannot be replicated autonomously in Coryneform
 PT bacteria - useful for isolating genes involved in amino acid
 PT biosynthesis
 XX
 PS Claim 3; Page 13-15; 16pp; Japanese.
 XX
 CC A plasmid which cannot be replicated autonomously in a Coryneform
 CC bacterium carries a DNA region having at least one inverted repeat
 CC (IR) present upstream or downstream of an open reading frame within
 CC an inserted sequence. The IRs are derived from the 5' upstream and
 CC 3' downstream regions of a marker gene derived from a Coryneform
 CC bacterium. The plasmid is used in a method for obtaining a mutant in
 CC which a Coryneform bacterium is transformed with the plasmid and the
 CC strain expressing the marker gene is separated by using the marker gene
 CC of the transformant as the index. Genes relating to amino acid
 CC biosynthesis can be easily isolated by using the mutant. The present
 CC sequence is that of a protein encoded by the C. glutamicum open
 CC reading frame.
 XX
 SQ Sequence 436 AA;
 Query Match 42.1%; Score 51; DB 17; Length 436;
 Best Local Similarity 41.2%; Pred. No. 10;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 3 HHIFRGIVHGKTIHKL 19
 DB 417 HYILRLIHSGQLVHKL 433
 RESULT 11
 AAW23219
 ID AAW23219 standard; Protein; 436 AA.
 XX
 AC AAW23219;

```

XX 10-NOV-1997 (first entry)
DT
XX
DE Protein product of B. lactofermentum insertion sequence IS719.
XX
KM Insertion sequence; IS719; transposable element; Coryneform;
KW bacterium; bacteria; amplification; artificial transposon;
XX inverted repeat; amino acid; biosynthesis; aspartokinase;
XX dihydrodipicolinic acid; synthetase; synthase.
OS
XX Brevibacterium lactofermentum.
XX
XX EP756007-A2.
XX
XX 29-JAN-1997.
XX
XX 28-JUN-1996; 96EP-0110491.
XX
XX 30-JUN-1995; 95JP-0166541.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Hayakawa A, Hirano S, Izui M, Matsui H, Moriya M,
PI Sugimoto M, Yokozeki K;
XX
XX WPI; 1997-10186/10.
XX
XX N-PSDB; AAT79580.
XX
XX Amplifying gene, esp. aspartokinase, in Coryneform chromosome using
PT artificial transposon - useful for amino acid synthesis, esp. lysine
XX
XX Disclosure; Pages 42-47; 122pp; English.
PS
XX
XX The present sequence is the protein product of the Brevibacterium
CC lactofermentum strain AJ12036 insertion sequence, IS719, i.e. a
CC transposable element derived from a Coryneform bacterium.
CC IS719 can be used in a novel method for the amplification of a
CC desired gene, comprising the formation of an artificial transposon
CC having a structure such that a drug resistance gene and the
CC desired gene are held between inverted repeats. The artificial
CC transposon, which is transposable in a Coryneform bacterium, is
CC transduced into a Coryneform bacterium, transposed into its
CC chromosome and the desired gene transduced and amplified in the
CC chromosome. A transposase gene is preferably held between the
CC inverted repeats, which are derived from an insertion sequence of
CC a Coryneform bacterium, i.e. IS719. The drug resistance gene is
CC the chloramphenicol or tetracycline resistance gene, and the
CC desired gene is one that participates in amino acid biosynthesis,
CC especially an aspartokinase and/or dihydrodipicolinic acid synthetase
CC gene.
XX
XX Sequence 436 AA;
SQ
XX
XX Query Match 42.1%; Score 51; DB 18; Length 436;
XX Best Local Similarity 41.2%; Pred. No. 10;
XX Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 3 HHIFRGIVHVGKTIHKL 19
DB 417 HYLRLCLHSGQLVHKI 433

```

```

KM organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KIOM ) KIOMA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX
XX N-PSDB; AAH66152.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 4687; 246pp + Sequence listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 436 AA;
SQ
XX
XX Query Match 42.1%; Score 51; DB 22; Length 436;
XX Best Local Similarity 41.2%; Pred. No. 10;
XX Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 3 HHIFRGIVHVGKTIHKL 19
DB 417 HYLRLCLHSGQLVHKI 433

```

```

RESULT 12
AAG90933
ID AAG90933 standard; Protein; 436 AA.
XX
XX AAG90933;
AC
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum protein fragment SEQ ID NO: 4687.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW

```

```

RESULT 13
AAG92342
ID AAG92342 standard; Protein; 436 AA.
XX
XX AAG92342;
AC
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum protein fragment SEQ ID NO: 6096.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX

```

```

XX      18-DEC-2000; 2000EP-0127688.
PR      16-DEC-1999;   99JP-0377484.
PR      07-APR-2000; 2000JP-0159162.
PR      03-AUG-2000; 2000JP-0280988.
XX      (KYOW ) KYOMA HAKKO KOGYO KK.
PA      Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI      Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX      WPI; 2001-376931/40.
DR      N-PDSB; AAH67561.
XX
XX      Novel polynucleotides derived from Coryneform bacteria, for identifying
PT      mutation point of a gene, measuring expression of a gene, analysing
PT      expression profile or pattern of a gene and identifying homologous gene
PT      _
XX
PS      Claim 17; SEQ ID NO: 6096; 246pp + Sequence Listing; English.
XX
XX      The present invention provides a number of nucleotide and protein
CC      sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC      are useful for identifying the mutation point of a gene derived from a
CC      mutant of coryneform bacterium, measuring expression amount and
CC      analysing the expression profile or expression pattern of a gene derived
CC      from Coryneform bacterium, and identifying a homologue of a gene derived
CC      from coryneform bacterium. Coryneform bacteria are useful for producing
CC      amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC      particularly L-lysine. The present sequence is a protein described
CC      in the exemplification of the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from the
CC      European Patent Office.
XX
SQ      Sequence    436 AA;
XX
XX      Query Match          42.1%; Score 51; DB 22; Length 436;
XX      Best Local Similarity 41.2%; Pred. No. 10;
XX      Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
OY      3 HHIFRGIVHGKTIHTL 19
        ||:||:||:||:||:
DB      417 HYLRLCHSGQLVHKI 433
RESUT 14
AAG92467
ID      AAG92467 standard; Protein; 436 AA.
AC      AAG92467;
XX
DT      26-SEP-2001 (first entry)
DE      C glutamicum protein fragment SEQ ID NO: 6221.
XX
XX      Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM      organic acid synthesis.
OS      Corynebacterium glutamicum.
PN      EP1108790-A2.
PD      20-JUN-2001.
PF      18-DEC-2000; 2000EP-0127688.
XX
PR      16-DEC-1999;   99JP-0377484.
PR      07-APR-2000; 2000JP-0159162.
PR      03-AUG-2000; 2000JP-0280988.
XX
PA      (KYOW ) KYOMA HAKKO KOGYO KK.
XX

```

PI	Nakagawa S, Mitoiguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	WP1, 2001-376931/40.
DR	N-PSDB; AAH67686.
XX	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
PT	-
PS	Claim 17; SEQ ID NO: 6221; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a protein described
CC	in the exemplification of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
CC	
SQ	Sequence 436 AA;
Query March	42.1%; Score 51; DB 22; Length 436;
Best Local Similarity	41.2%; Pred. No. 10;
Matches 7; Conservative	6; Mismatches 4; Indels 0; Gaps 0;
OY	3 HHIFGIIVHVGKTIHKL 19
	: : : : : : : :
Db	417 HYIKRLIHSGIVHKK 433
RESULT 15	
ID	AAU60341
AAU60341	standard; Protein; 78 AA.
XX	
AC	AAU60341;
XX	
DT	27-FEB-2002 (first entry)
XX	
DE	Propionibacterium acnes immunogenic protein #21237.
XX	
KW	SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW	uvulitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW	inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW	dermatological; osteopathic; neuroprotectant.
XX	
OS	Propionibacterium acnes.
XX	
PN	WO200181581-A2.
XX	
PD	01-NOV-2001.
XX	
PF	20-APR-2001; 2001WO-US12865.
XX	
PR	21-APR-2000; 2000US-199047P.
PR	02-JUN-2000; 2000US-208841P.
PR	07-JUL-2000; 2000US-216747P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX	WP1; 2001-616774/71.
DR	N-PSDB; AAS59609.
XX	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:54:25 ; Search time 11.5 Seconds
(without alignments)
56.287 Million cell updates/sec

Title: US-09-929-788-3
Perfect score: 121
Sequence: 1 FFHHIERGIVHVKTIHKLVTVG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/iaa/backflist.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	42.1	436	1	US-08-674-168-6
2	47	38.8	436	1	US-08-146-010A-2
3	47	38.8	436	1	US-08-674-168-2
4	43	35.5	399	2	US-08-834-655-4
5	43	35.5	399	3	US-08-834-033A-4
6	43	35.5	399	4	US-09-363-574-4
7	43	35.5	399	4	US-09-363-526-4
8	43	35.5	399	4	US-09-330-235-20
9	42	34.7	466	4	US-09-134-001C-3526
10	41.5	34.3	23	2	US-08-808-277A-18
11	41	33.9	115	4	US-09-374-135-3
12	41	33.9	471	4	US-09-134-001C-4904
13	41	33.9	865	1	US-07-803-633A-13
14	40	33.1	223	4	US-08-147-772-2
15	40	33.1	223	4	US-08-147-772-2
16	40	33.1	288	2	US-08-456-104-6
17	40	33.1	288	2	US-08-101-624-23
18	40	33.1	288	2	US-08-751-767A-6
19	40	33.1	288	2	US-08-153-262-2
20	40	33.1	288	3	US-08-479-744A-29
21	40	33.1	288	4	US-08-280-757B-29
22	40	33.1	288	4	US-09-159-135-2
23	40	33.1	288	4	US-08-205-697A-19
24	40	33.1	288	4	US-08-702-525-19
25	40	33.1	288	4	US-09-450-798-2
26	40	33.1	288	4	US-08-403-253A-2
27	40	33.1	288	4	US-09-651-200-13

28	40	33.1	288	4	US-09-651-200-14	Sequence 14, Appl
29	40	33.1	288	5	PCT-US95-02576-19	Sequence 19, Appl
30	40	33.1	345	2	US-08-758-621-14	Sequence 14, Appl
31	40	33.1	345	4	US-09-107-858-14	Sequence 14, Appl
32	40	33.1	473	4	US-09-171-945-131	Sequence 131, Appl
33	40	33.1	810	2	US-08-820-170A-34	Sequence 34, Appl
34	40	33.1	810	3	US-09-055-689-34	Sequence 34, Appl
35	40	33.1	810	4	US-09-273-565-34	Sequence 34, Appl
36	40	33.1	810	4	US-09-565-538-34	Sequence 34, Appl
37	40	33.1	810	4	US-09-661-468-34	Sequence 34, Appl
38	39	32.2	270	4	US-09-693-147-4	Sequence 4251, Ap
39	39	32.2	434	4	US-09-134-001C-4251	Sequence 2, Appl
40	39	32.2	768	2	US-08-560-398-2	Sequence 8, Appl
41	39	32.2	885	1	US-08-042-747A-8	Sequence 23, Appl
42	39	32.2	885	3	US-08-804-439A-23	Sequence 23, Appl
43	39	32.2	885	3	US-08-720-229-23	Sequence 4, Appl
44	39	32.2	891	1	US-08-042-747A-6	Sequence 6, Appl
45	38.5	31.8	668	1	US-08-205-018-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-674-168-6
; Sequence 6, Application US/08674168
; Patent No. 5804414
; GENERAL INFORMATION:
; APPLICANT: MORIYA, Mika
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: YOKOZAKI, Kenzo
; APPLICANT: HIRANO, Seiko
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: IZUI, Masako
; APPLICANT: SUGIMOTO, Masakazu
; TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
; TITLE OF INVENTION: ARTIFICIAL TRANSPOSON
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P. C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,168
; FILING DATE: 01-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-166541
; FILING DATE: 30-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-810-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-674-168-6

Query Match	42.1%;	Score 51;	DB 1;	Length 436;
Best Local Similarity	41.2%;	Pred. No. 2.9;		
Matches	7;	Conservative	6;	Mismatches 4;
				Indels

```
QY      3 HHIFRGIVHVGKTIHKL 19
         ||:||:||:||:
Db     417 HYLRLIHSGLVHKI 433
```

RESULT 2
US-08-146-010A-2

Query Match	38.8%;	Score 47;	DB 1;	Length 436;
Best Local Similarity	41.2%;	Pred. No. 12;		
Matches	7;	Conservative	5;	Mismatches 5; Indels 0; Gaps 0

QY 3 HHIFRGIVHVGKTIHKL 19
|:|:|:|:|:|:
Db 417 HYLRCLIHSGQLTHKI 433

US-08-674-168-2
RESULT 3
; Sequence 2, Application US/08674168
; Patent No. 5804414
; GENERAL INFORMATION:
; APPLICANT: MORIYA, Mika
; APPLICANT: MATSUZ, Hiroshi
; APPLICANT: YOKOZAKI, Kenzo

APPLICANT: HIRANO, Seiko
 APPLICANT: HAYAKAWA, Atsushi
 APPLICANT: IZUI, Masako
 APPLICANT: SUGIMOTO, Masakazu
 TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
 TITLE OF INVENTION: ARTIFICIAL TRANSPOSON
 NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SHIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P. C.
 STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
 CITY: ARLINGTON
 STATE: VIRGINIA
 COUNTRY: USA
 ZIB: 33203

Query Match	38.8%;	Score 47;	DB 1;	Length 436;
Best Local Similarity	41.2%;	Pred. No. 12;		
Matches	7;	Conservative	5;	Mismatches 5;
				Indels 0;
				Gaps 0

Qy	3	HHIFRGIVHGKTIHKL	19
		: : : :	
Db	417	HYLRCLHSGQLTHKI	433

RESULT 4
 US-08-834-655-4
 : Sequence 4, Application US/08834655
 : Patent No. 5968809
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: KNUTZON, DEBORAH
 : APPLICANT: MORKERJI, PRADIP
 : APPLICANT: HUANG, YUNG-SHENG
 : APPLICANT: THURMOND, JENNIFER
 : APPLICANT: CHAUDHARY, SUNITA
 :
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS
 : TITLE OF INVENTION: OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS IN PLANTS
 :
 : NUMBER OF SEQUENCES: 18
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSER: RAE-VENTER LAW GROUP, P.C.
 : STREET: 260 SHERIDAN AVENUE, P.O. BOX 60039
 : CITY: PALO ALTO
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94306
 :
 : COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,655
FILING DATE: 11-Apr-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER, BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGAB.124.00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 328-4400
TELEFAX: (650) 328-4477
TELEX: N/A
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-834-655-4

Query Match 35.5% Score 43; DB 2; Length 399;
Best Local Similarity 47.1%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVKTIH 17
| : | | | | |
Db 326 FLDDHMFHGIHTVHAH 342

RESULT 5
US-08-834-033A-4
Sequence 4, Application US/08834033A
Patent No. 6075183
GENERAL INFORMATION:
APPLICANT: KNUTZON, DEBORAH
APPLICANT: MURKERJI, PRADIP
APPLICANT: HUANG, YUNG-SHENG
APPLICANT: THURMOND, JENNIFER
APPLICANT: CHAUDHARY, SUNITA
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS
OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS IN PLANTS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LIMBACH AND LIMBACH, L.L.P.
STREET: 2001 FERRY BUILDING
CITY: SAN FRANCISCO
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,033A
FILING DATE: 11-Apr-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: WARD, MICHAEL R.
REGISTRATION NUMBER: 38,651
REFERENCE/DOCKET NUMBER: CGAB-300.USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 433-4150
TELEFAX: (415) 433-8716
TELEX: N/A
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-834-033A-4

Query Match 35.5% Score 43; DB 3; Length 399;
Best Local Similarity 47.1%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVKTIH 17
| : | | | | |
Db 326 FLDDHMFHGIHTVHAH 342

RESULT 6
US-09-363-574-4
Sequence 4, Application US/09363574
Patent No. 6136574
GENERAL INFORMATION:
APPLICANT: KNUTZON, DEBORAH
APPLICANT: MURKERJI, PRADIP
APPLICANT: HUANG, YUNG-SHENG
APPLICANT: THURMOND, JENNIFER
APPLICANT: CHAUDHARY, SUNITA
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS
OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS IN PLANTS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: LIMBACH AND LIMBACH L.L.P.
STREET: 2001 FERRY BUILDING
CITY: SAN FRANCISCO
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/363,574
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WARD, MICHAEL R.
REGISTRATION NUMBER: 38,651
REFERENCE/DOCKET NUMBER: CGAB-202 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 433-4150
TELEFAX: (415) 433-8716
TELEX: N/A
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-363-574-4

Query Match 35.5% Score 43; DB 4; Length 399;
Best Local Similarity 47.1%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVKTIH 17
| : | | | | |
Db 326 FLDDHMFHGIHTVHAH 342

RESULT 7
US-09-363-526-4

Sequence 4, Application US/09363526
Patent No. 6410288
GENERAL INFORMATION:
APPLICANT: KUTZON, DEBORAH
APPLICANT: MURKERT, PRADIP
APPLICANT: HUANG, YUNG-SHENG
APPLICANT: THURMOND, JENNIFER
APPLICANT: CHAUDHARY, SUNITA
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS
TITLE OF INVENTION: OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS IN PLANTS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: LIMBACH AND LIMBACH L.L.P.
STREET: 2001 FERRY BUILDING
CITY: SAN FRANCISCO
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/363,526
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WARD, MICHAEL R.
REGISTRATION/DOCKET NUMBER: 38,651
REFERENCE/DOCKET NUMBER: CGAB-201 USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 433-4150
TELEFAX: (415) 433-8716
TELEX: N/A
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-363-526-4

Query Match 35.5%; Score 43; DB 4; Length 399;
Best Local Similarity 47.1%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHFRGIVHVGKTIH 17
DB 326 FLDDHMFHGIHTVVAHH 342

RESULT 8
US-09-330-235-20
Sequence 20, Application US/09330235
Patent No. 6459818
GENERAL INFORMATION:
APPLICANT: Knutzon, Debbie
TITLE OF INVENTION: POLYUNSATURATED FATTY ACIDS IN PLANTS
FILE REFERENCE: MOCO.156.00US
CURRENT APPLICATION NUMBER: US/09/330,235
CURRENT FILING DATE: 1999-06-10
PRIOR APPLICATION NUMBER: 60/089,043
PRIOR FILING DATE: 1998-06-12
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 20
LENGTH: 399
TYPE: PRT
ORGANISM: Mortierella alpina
US-09-330-235-20

Query Match 35.5%; Score 43; DB 4; Length 399;
Best Local Similarity 47.1%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHFRGIVHVGKTIH 17
DB 326 FLDDHMFHGIHTVVAHH 342

RESULT 9
US-09-134-001C-3526
Sequence 3526, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3526
LENGTH: 466
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3526

Query Match 34.7%; Score 42; DB 4; Length 466;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 2 FHHFRGIVHVGKTIHKLVTG 22
DB 47 FHYHDMSTVNAIVSLTSLTG 67

RESULT 10
US-08-808-277A-18
Sequence 18, Application US/08808277A
Patent No. 5998374
GENERAL INFORMATION:
APPLICANT: LEHRER, ROBERT I.
APPLICANT: LEE, IN-HEE
APPLICANT: ZHAO, CHENGQUAN
TITLE OF INVENTION: CLAVALSPIRINS
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,277A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 220002057300

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
TELEFAX: 202-822-0168
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5998374e
US-08-808-277A-16

Query Match 34.3%; Score 41.5; DB 2; Length 23;
Best Local Similarity 36.8%; Pred. No. 4.1;
Matches 7; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

Qy 2 FHHIFRGIVH-VGKTHKL 19
Db 1 FFGIGGIVHGVGRVHKL 19

RESULT 11
US-09-374-135-3
Sequence 3, Application US/09374135
Patent No. 6277972
GENERAL INFORMATION:
APPLICANT: Afari, Daniel E.
APPLICANT: Hubert, Rene S.
APPLICANT: Leong, Arthur B.
APPLICANT: Saffran, Douglas C.
APPLICANT: Jakobovits, Aya
TITLE OF INVENTION: BPC-1: A SECRETED BRAIN-SPECIFIC PROTEIN EXPRESSED AND
FILE REFERENCE: 1703-017.US1
CURRENT APPLICATION NUMBER: US/09/374,135
PRIOR FILING DATE: 1999-08-10
PRIOR FILING DATE: 1998-08-10
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 115
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-374-135-3

Query Match 33.9%; Score 41; DB 4; Length 115;
Best Local Similarity 38.1%; Pred. No. 25;
Matches 8; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

Qy 2 FHHIFRGIVHGVKTHKLVTG 22
Db 33 FHHVF---HIESTYDKIDAG 49

RESULT 12
US-09-134-001C-4904
Sequence 4904, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4904
LENGTH: 471
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4904

Query Match 33.9%; Score 41; DB 4; Length 471;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 IFRGIVHGVK 14
Db 361 IYRGIVHGR 370

RESULT 13
US-07-803-633A-13
Sequence 13, Application US/07803633A
Patent No. 5369025
GENERAL INFORMATION:
APPLICANT: NAZERIAN, Keyvan
APPLICANT: LEE, Lucy F.
APPLICANT: VANAGIDA, No. 5369025oru
APPLICANT: OGAWA, Ryohei
APPLICANT: LI, Yi
TITLE OF INVENTION: RECOMBINANT FOWLPOX VACCINE FOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 301 No. 5369025th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/803,633A
FILING DATE: 19911210
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1644-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 241-1300
TELEFAX: (703) 241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 865 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-803-633A-13

Query Match 33.9%; Score 41; DB 1; Length 865;
Best Local Similarity 44.4%; Pred. No. 2e+02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 5 IFRGIVHGVKTHKLVTG 22
Db 695 IFNGMGVGAIGKVVG 712

RESULT 14
US-08-961-083-162
Sequence 162, Application US/08961083
Patent No. 6159469

```

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961.083
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
FILING DATE:
NAME: Brooks, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-162

Query Match 33.1%; Score 40; DB 4; Length 223;
Best Local Similarity 27.6%; Pred. No. 70;
Matches 8; Conservative 5; Mismatches 8; Indels 8; Gaps 1;

QY 2 FHIIIFRGIVH-----GTHHKLVYTG 22
Db 81 YHILGMGLAHVDGTPPLDGTGIRSVIAG 109

RESULT 15
US-08-147-772-2
Sequence 2, Application US/08147772
Patent No. 5858776
GENERAL INFORMATION:
APPLICANT: Ostrand-Rosenberg, Suzanne
APPLICANT: Basker, Sivasubramanian
APPLICANT: Glimcher, Laurie H.
APPLICANT: Freeman, Gordon J.
TITLE OF INVENTION: Tumor Cells With Increased Immunogenicity
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/147,772
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPT-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: B cell activation antigen; natural ligand
DESCRIPTION: for CD28 T cell surface antigen; transmembrane protein
FEATURE:
NAME/KEY: signal sequence
LOCATION: -34 to -1
IDENTIFICATION METHOD: amino terminal sequencing of
IDENTIFICATION METHOD: soluble protein
OTHER INFORMATION: hydrophobic
FEATURE:
NAME/KEY: extracellular domain
LOCATION: 1 to 208
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: transmembrane domain
LOCATION: 209 to 235
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: intracellular domain
LOCATION: 236 to 254
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 19 to 21
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 55 to 57
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 64 to 66
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 152 to 154
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 173 to 175
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 177 to 179
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
```

```
NAME/KEY: N-linked glycosylation
LOCATION: 192 to 194
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: N-linked glycosylation
LOCATION: 198 to 200
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: Ig V-set domain
LOCATION: 1 to 104
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
FEATURE:
NAME/KEY: Ig C-set domain
LOCATION: 105 to 202
IDENTIFICATION METHOD: similarity with known
IDENTIFICATION METHOD: sequence
PUBLICATION INFORMATION:
AUTHORS: FREEDMAN, GORDON J.
AUTHORS: FREEDMAN, ARNOLD S.
AUTHORS: SEGIL, JEFFREY M.
AUTHORS: LEE, GRACE
AUTHORS: WHITMAN, JAMES F.
AUTHORS: NADLER, LEE M.
TITLE: B7, A New Member Of The Ig Superfamily With
JOURNAL: The Journal of Immunology
VOLUME: 143
ISSUE: 8
PAGES: 2714-2722
DATE: 15-OCT-1989
RELEVANT RESIDUES IN SEQ ID NO: 2: From -26 to 262
US-08-147-772-2
```

```
Query Match 33.1%; Score 40; DB 2; Length 288;
Best Local Similarity 33.3%; Pred. No. 91;
Matches 6; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
QY 4 HIFRGIVHGKTIHKLVT 21
Db 30 HFCGVIHVTKEKEVAT 47
```

Search completed: May 23, 2003, 08:57:46
Job time : 11.5 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:56:50 ; Search time 16 Seconds

(without alignments)
136.362 Million cell updates/sec

Title: US-09-929-788-3

Perfect score: 121
Sequence: 1 FFHHIFRGIVHVGKTIHKLVTG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	22	9	US-09-929-788-3
2	118	97.5	22	9	US-09-929-788-2
3	115	95.0	23	9	US-10-076-816-42
4	84	69.4	22	9	US-09-929-788-1
5	51	42.1	436	9	US-09-738-626-4687
6	51	42.1	436	9	US-09-738-626-6096
7	46	38.0	44	9	US-09-738-626-6221
8	46	38.0	384	9	US-09-929-788-4
9	46	38.0	118	10	US-09-738-626-6407
10	45.5	37.6	392	9	US-09-867-550-198
11	45.5	37.6	392	9	US-09-912-935-23
12	45.5	37.6	392	9	US-09-912-935-25
13	45.5	37.6	425	9	US-09-912-935-35
14	45.5	37.6	449	9	US-09-912-935-34
15	45.5	37.6	499	9	US-09-912-935-31
16	45.5	37.6	529	9	US-10-066-500-128
17	45.5	37.6	529	9	US-10-174-590-472
18	45.5	37.6	529	9	US-10-176-987-472
19	45.5	37.6	529	9	US-10-176-987-472

20	45.5	37.6	529	9	US-09-912-935-28	Sequence 28, Appl
21	45.5	37.6	529	9	US-09-912-935-40	Sequence 40, Appl
22	45.5	37.6	529	9	US-10-173-706-472	Sequence 472, App
23	45.5	37.6	529	9	US-10-175-738-472	Sequence 472, App
24	45.5	37.6	529	9	US-10-175-752-472	Sequence 472, App
25	45.5	37.6	529	9	US-10-176-482-472	Sequence 472, App
26	45.5	37.6	529	9	US-10-176-757-472	Sequence 472, App
27	45.5	37.6	529	9	US-10-176-913-472	Sequence 472, App
28	45.5	37.6	529	9	US-10-180-552-472	Sequence 472, App
29	45.5	37.6	529	9	US-10-180-557-472	Sequence 472, App
30	45.5	37.6	529	9	US-10-173-700-472	Sequence 472, App
31	45.5	37.6	529	9	US-10-174-579-472	Sequence 472, App
32	45.5	37.6	529	9	US-10-174-582-472	Sequence 472, App
33	45.5	37.6	529	9	US-10-174-588-472	Sequence 472, App
34	45.5	37.6	529	9	US-10-175-739-472	Sequence 472, App
35	45.5	37.6	529	9	US-10-175-740-472	Sequence 472, App
36	45.5	37.6	529	9	US-10-175-743-472	Sequence 472, App
37	45.5	37.6	529	9	US-10-176-488-472	Sequence 472, App
38	45.5	37.6	529	9	US-10-176-747-472	Sequence 472, App
39	45.5	37.6	529	9	US-10-176-750-472	Sequence 472, App
40	45.5	37.6	529	9	US-10-176-750-472	Sequence 472, App
41	45.5	37.6	529	9	US-10-176-985-472	Sequence 472, App
42	45.5	37.6	529	9	US-10-176-987-472	Sequence 472, App
43	45.5	37.6	529	9	US-10-176-991-472	Sequence 472, App
44	45.5	37.6	529	9	US-10-176-992-472	Sequence 472, App
45	45.5	37.6	529	9	US-10-176-992-472	Sequence 472, App

ALIGNMENTS

```
RESULT 1
US-09-929-788-3
; Sequence 3, Application US/09929788
; Publication No. US20030083247A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030083247A1a, Edward J.
; APPLICANT: Silphaduang, Umaporn
; TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
; FILE REFERENCE: 5051.519
; CURRENT APPLICATION NUMBER: US/09/929,788
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,354
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1.1
; SEQ ID NO 3
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Morone saxatilis x Morone chrysops
US-09-929-788-3

Query Match      100.0%; Score 121, DB 9, Length 22;
Best Local Similarity 100.0%; Pred. No. 5, Se-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FFHHIFRGIVHVGKTIHKLVTG 22
Db      1 FFHHIFRGIVHVGKTIHKLVTG 22

RESULT 2
US-09-929-788-2
; Sequence 2, Application US/09929788
; Publication No. US20030083247A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030083247A1a, Edward J.
; APPLICANT: Silphaduang, Umaporn
; TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
; FILE REFERENCE: 5051.519
; CURRENT APPLICATION NUMBER: US/09/929,788
; PRIOR FILING DATE: 2001-08-14
; SOFTWARE: PatentIn version 3.1.1
; SEQ ID NO 2
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Morone saxatilis x Morone chrysops
```

```

; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Morone saxatilis x Morone chrysops
US-09-929-788-2

```

```

Query Match          97.5%; Score 118; DB 9; Length 22;
Best Local Similarity 95.5%; Pred. No. 1.6e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 FFHHIFRGIVHVKTIHKLVTG 22
Db 1 FFHHIFRGIVHVKTIHKLVTG 22

```

```

RESULT 3
US-10-076-816-42
; Sequence 42, Application US/10076816
; Publication No. US20030056244A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Ning
; APPLICANT: Rodriguez, Raymond
; APPLICANT: Hagie, Frank E.
; TITLE OF INVENTION: Feed Additive Compositions and Methods
; FILE REFERENCE: 50665-8021.0S00
; CURRENT APPLICATION NUMBER: US/10/076,816
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/269,188
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/847,232
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/266,929
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: US 60/201,182
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Morone sp.
US-10-076-816-42

```

```

Query Match          95.0%; Score 115; DB 9; Length 23;
Best Local Similarity 95.5%; Pred. No. 4.7e-11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 FFHHIFRGIVHVKTIHKLVTG 22
Db 1 FFHHIFRGIVHVKTIHKLVTG 22

```

```

RESULT 4
US-09-929-788-1
; Sequence 1, Application US/09929788
; Publication No. US2003008347A1
; GENERAL INFORMATION:
; APPLICANT: No. US2003008347A1a, Edward J.
; APPLICANT: Silphaduang, Umaborn
; TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
; FILE REFERENCE: 5051.519
; CURRENT APPLICATION NUMBER: US/09/929,788
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,354
; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 22
; TYPE: PRT

```

```

; ORGANISM: Morone saxatilis x Morone chrysops
US-09-929-788-1

```

```

Query Match          69.4%; Score 84; DB 9; Length 22;
Best Local Similarity 63.6%; Pred. No. 2.3e-06;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 FFHHIFRGIVHVKTIHKLVTG 22
Db 1 FFHHIFRGIVHGRSIRGRFLTG 22

```

```

RESULT 5
US-09-738-626-4687
; Sequence 4687, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4687
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4687

```

```

Query Match          42.1%; Score 51; DB 9; Length 436;
Best Local Similarity 41.2%; Pred. No. 6.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 HHIFRGIVHVKTIHKL 19
Db 417 HYLRLIHSGQLVHKI 433

```

```

RESULT 6
US-09-738-626-6096
; Sequence 6096, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18

```

```
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6096
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6096
```

```
Query Match          42.1%; Score 51; DB 9; Length 436;
Best Local Similarity 41.2%; Pred. No. 6.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 3 HHFRGIVHVGKTIHKL 19
Db 417 HYLRCLHSGQLVHKI 433
```

```
RESULT 7
US-09-738-626-6221
; Sequence 6221, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKOHO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6221
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6221
```

```
Query Match          42.1%; Score 51; DB 9; Length 436;
Best Local Similarity 41.2%; Pred. No. 6.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 3 HHFRGIVHVGKTIHKL 19
Db 417 HYLRCLHSGQLVHKI 433
```

```
RESULT 8
US-09-929-788-4
; Sequence 4, Application US/09929788
; Publication No. US20030083247A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030083247A1a, Edward J.
; APPLICANT: Silphadung Umaporn
; TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
```

```
; FILE REFERENCE: 5051, 519
; CURRENT APPLICATION NUMBER: US/09/929,788
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,354
; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Morone saxatilis x Morone chrysops
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (20)-(120)
; OTHER INFORMATION: "X" is unknown
US-09-929-788-4
```

```
Query Match          38.0%; Score 46; DB 9; Length 44;
Best Local Similarity 37.0%; Pred. No. 2.9;
Matches 10; Conservative 4; Mismatches 7; Indels 6; Gaps 1;
```

```
Oy 1 FFHHFRGIVHVGKTIHKLVT 21
Db 1 FFHHLFRGAAIFRGARQGXRAHKVVS 27
```

```
RESULT 9
US-09-738-626-6407
; Sequence 6407, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKOHO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6407
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6407
```

```
Query Match          38.0%; Score 46; DB 9; Length 384;
Best Local Similarity 41.2%; Pred. No. 3.3;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
```

```
Oy 1 FFHHFRGIVHVGKTIH 17
Db 325 FVHHVHRGWTLDATLH 341
```

```
RESULT 10
US-09-867-550-198
; Sequence 198, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
```

```

APPLICANT: Leach, Martin D.
APPLICANT: Mehraban, Foad,
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie
APPLICANT: Topper, James
TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
TITLE OF INVENTION: Thereby
FILE REFERENCE: 21402-013 (Cura-313)
CURRENT APPLICATION NUMBER: US/09/867,550
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: USSN 60/208,427
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 198
LENGTH: 118
TYPE: PRT
ORGANISM: Homo sapiens
US-09-867-550-198

Query Match      37.6%; Score 45.5; DB 10; Length 118;
Best Local Similarity 25.0%; Pred. No. 10;
Matches       7; Conservative   8; Mismatches    6; Indels   7; Gaps    1;

OY      1 FFHHIR-----GIHVGTKIHKLV 21
| | | | | : | | | | | : | | | | |
DB      53 FYGHFLREITVATGFGFIYTGEVVHRLT 80

RESULT 11
US-09-912-935-23
Sequence 23, Application US/09912935
Publication No. US20030022825A1
GENERAL INFORMATION:
APPLICANT: Nishikawa, Mitsuo et al.
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
FILE REFERENCE: 32066/37483
CURRENT APPLICATION NUMBER: US/09/912,935
CURRENT FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: PCT/US00/35260
PRIOR FILING DATE: 2000-12-23
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 23
LENGTH: 392
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-935-23

Query Match      37.6%; Score 45.5; DB 9; Length 392;
Best Local Similarity 25.0%; Pred. No. 41;
Matches       7; Conservative   8; Mismatches    6; Indels   7; Gaps    1;

OY      1 FFHHIR-----GIHVGTKIHKLV 21
| | | | | : | | | | | : | | | | |
DB      30 FYGHFLREITVATGFGFIYTGEVVHRLT 57

RESULT 12
US-09-912-935-25
Sequence 25, Application US/09912935
Publication No. US20030022825A1
GENERAL INFORMATION:
APPLICANT: Nishikawa, Mitsuo et al.
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
FILE REFERENCE: 32066/37483
CURRENT APPLICATION NUMBER: US/09/912,935
CURRENT FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: PCT/US00/35260
PRIOR FILING DATE: 2000-12-23
NUMBER OF SEQ ID NOS: 53
```

```

SOFTWARE: PatentIn version 3.0
SEQ ID NO 25
LENGTH: 392
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-935-25

Query Match          37.6%; Score 45.5; DB 9; Length 392;
Best Local Similarity 25.0%; Pred. No. 41;
Matches      7; Conservative      8; Mismatches      6; Indels      7; Gaps      1;

OY      1 FFHHIFR-----GIHVNGKTIHKLVLT 21
| : | | | | | | | | | | | | | | | | | | | |
Db      30 FYGHFLREITVATGCGFIYTGBVVHRMLT 57

RESULT 13
US-09-912-935-35
Sequence 35, Application US/09912935
Publication No. US20030022825A1
GENERAL INFORMATION:
APPLICANT: Nishikawa, Mitsuo et al.
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
FILE REFERENCE: 32066/37483
CURRENT APPLICATION NUMBER: US/09/912.935
CURRENT FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: PCT/US00/35260
PRIOR FILING DATE: 2000-12-23
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO 35
LENGTH: 425
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-935-35

Query Match          37.6%; Score 45.5; DB 9; Length 425;
Best Local Similarity 25.0%; Pred. No. 44;
Matches      7; Conservative      8; Mismatches      6; Indels      7; Gaps      1;

OY      1 FFHHIFR-----GIHVNGKTIHKLVLT 21
| : | | | | | | | | | | | | | | | | | | | |
Db      65 FYGHFLREITVATGCGFIYTGBVVHRMLT 92

RESULT 14
US-09-912-935-34
Sequence 34, Application US/09912935
Publication No. US20030022825A1
GENERAL INFORMATION:
APPLICANT: Nishikawa, Mitsuo et al.
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
FILE REFERENCE: 32066/37483
CURRENT APPLICATION NUMBER: US/09/912.935
CURRENT FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: PCT/US00/35260
PRIOR FILING DATE: 2000-12-23
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.0
SEQ ID NO 34
LENGTH: 449
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-935-34

Query Match          37.6%; Score 45.5; DB 9; Length 449;
Best Local Similarity 25.0%; Pred. No. 47;
Matches      7; Conservative      8; Mismatches      6; Indels      7; Gaps      1;

OY      1 FFHHIFR-----GIHVNGKTIHKLVLT 21
| : | | | | | | | | | | | | | | | | | | | |

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:53:20 ; Search time 14 Seconds

(without alignments)
151.068 Million cell updates/sec

Title: US-09-929-788-3

Perfect score: 121

Sequence: 1 FFHHIFRGIVHGKTIHKLVGTG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	42.1	436	2	S43613
2	48	39.7	351	2	AD0273
3	48	39.7	404	2	S64944
4	47	38.8	173	2	B69354
5	47	38.8	436	2	JC4742
6	47	38.8	476	2	G84634
7	47	38.8	612	2	AC1208
8	47	38.8	612	2	AF1564
9	46	38.0	907	2	H71031
10	46	38.0	907	2	B75182
11	45.5	37.6	174	2	G84600
12	45	37.2	201	2	T36329
13	44	36.4	311	2	E95088
14	44	36.4	311	2	H97955
15	44	36.4	513	2	C75553
16	44	36.4	905	1	S25564
17	44	36.4	952	2	C82309
18	44	36.4	1213	2	E69255
19	44	36.4	2550	2	B53435
20	43.5	36.0	391	2	C87615
21	43	35.5	192	2	A42116
22	43	35.5	223	2	T21669
23	43	35.5	223	2	T47003
24	43	35.5	299	2	AF0238
25	43	35.5	326	2	B40241
26	43	35.5	330	2	S26596
27	43	35.5	372	2	T52621
28	43	35.5	415	2	AB0573
29	43	35.5	523	2	T26740

30	43	35.5	599	2	AC1524	metyl-accepting ch
31	43	35.5	601	2	AC1165	metyl-accepting ch
32	43	35.5	774	2	A13372	malate dehydrogena
33	42.5	35.1	878	2	B84977	alanine-tRNA ligas
34	42.5	35.1	1068	2	T04112	pol protein homolo
35	42	34.7	82	2	T31574	hypothetical prote
36	42	34.7	194	2	B75613	conserved hypothet
37	42	34.7	227	2	B97624	hypothetical prote
38	42	34.7	368	2	E82645	NAD(P)H-dependent
39	42	34.7	403	2	AF2151	aminotransferase l
40	42	34.7	474	2	S18452	variant surface gl
41	42	34.7	478	2	E90495	aldehyde dehydroge
42	42	34.7	478	2	A75520	pyrimidine-nucleos
43	42	34.7	480	2	JC7812	BCR6 homologous zi
44	42	34.7	531	2	T11074	NADH2 dehydrogenas
45	42	34.7	544	2	E95412	hypothetical prote

ALIGNMENTS

RESULT 1

S43613 transposase (insertion sequence IS31831) - Corynebacterium glutamicum

C/Species: Corynebacterium glutamicum

C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999

C/Accession: S43613

R/Veres, A.A.; Inui, M.; Kobayashi, M.; Kurusu, Y.; Yokawa, H.

Mol. Microbiol. 11, 739-746, 1994

A/Title: Isolation and characterization of IS31831, a transposable element from Coryneba

A/Reference number: S43613; MUID:94254729; PMID:8196545

A/Accession: S43613

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-436 <VER>

A/Cross-references: EMBL: D17429; NID: g473697; PIDN: BAA04250.1; PID: d1004769; PID: g790952

Query Match 42.1%; Score 51; DB 2; Length 436;

Best Local Similarity 41.2%; Pred. No. 4.3;

Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 417 HYLRCLHSGQLVHKI 433

Query 3 HHIFRGIVHGKTIHKL 19

Db 417 HYLRCLHSGQLVHKI 433

RESULT 2

AD0273 Probable integral membrane protein YPO2242 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C/Accession: AD0273

R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AD0273

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-351 <KDR>

A/Cross-references: GB: AU590842; PIDN: CAC91048.1; PID: g15980242; GSPDB: GN00175

C/Genetics: YPO2242

Query Match 39.7%; Score 48; DB 2; Length 351;

Best Local Similarity 50.0%; Pred. No. 9.9;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 5 IFRGIVHGKTIHKLVGTG 22

Db 160 IFTGYTHSGENIHQLQVG 177

RESULT 3
 S64944
 hypothetical protein YLR107w - yeast (*Saccharomyces cerevisiae*)
 N/Alternate names: hypothetical protein L2904
 C/Species: *Saccharomyces cerevisiae*
 C/Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 19-Apr-2002
 C/Accession: S64944; S69394
 R/Verhasselt, P.; Voet, M.; Voelckaert, G.
 submitted to the Protein Sequence Database, May 1996
 A/Reference number: S64943
 A/Accession: S64944
 A/Molecule type: DNA
 A/Residues: 1-404 <VER>
 A/Cross-references: EMBL: Z73279; NID: g1360495; PIDN: CAA97672.1; PID: e245554; PID: g1360494
 A/Experimental source: strain S288C
 R/Verhasselt, P.; Voelckaert, G.
 submitted to the EMBL Data Library, September 1995
 A/Reference number: S69393
 A/Accession: S69394
 A/Molecule type: DNA
 A/Residues: 1-404 <VEW>
 A/Cross-references: EMBL: X89514; NID: g1297019; PIDN: CAA61685.1; PID: e198740; PID: g1297020
 C/Genetics:
 A/Gene: SGD: REX3
 A/Cross-references: SGD: S0004097
 A/Map position: 12R

```

Query Match          39.7%;   Score 48;   DB 2;   Length 404;
Beet Local Similarity 87.5%;   Pred. NO. 12;
Matches 7;   Conservative 1;   Mismatches 0;   Indels 0;   Gaps 0;

Cy      1  FFHHIFRG 8
      ||||:||||
      ||||:||||
Db      212  FFHHVFRG 219

RESULT 4
B69354
ferritin homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Aug-1999
C:Accession: B69354
R:Klikn, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaee
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: B69354
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <KLE>
A:Cross-References: GB:AE001047; GB:AE000782; NID:g2689370; PIDN:AA890406.1; PID:g2649777
C:Superfamily: ferritin

```

```

Query Match Similarity      38.8%; Score 47; DB 2; Length 173;
Best Local Similarity      47.4%; Pred. No. 6.4;
Matches      9; Conservative      4; Mismatches      6; Indels      0; Gaps      0;

Qy      2 FHHIFRGIVHVGKTIHKLV 20
      |::|::|::|::|
Db      89 FEHVEHEVNVTKRHHELV 107

RESULT 5
JC4742
transposase - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:IDaG: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 17-Mar-1999

```

C:Accession: J04742
 C:Correla. A.; Pisabarro, A.; Castro, J.M.; Martin, J.F.
 Gene 170, 91-94, 1996
 A:Title: Cloning and characterization of an IS-like element present in the genome of *Brevibacterium lactofermentum*
 A:Reference number: J04742; WUID:96200862; PMID:8621097
 A:Accession: J04742
 A:Molecule type: DNA
 A:Residues: 1-436 <COR>
 A:Cross-references: EMBL:266534
 A:Experimental source: ATCC 13869
 A:Note: The authors translated the initiation codon TGT for residue 1 as Val
 A:Note: The authors translated the codon ATT for residue 155 as Tyr
 A:Note: the source is designated as *Brevibacterium lactofermentum*
 C:Genetics:
 A:Gene: GTG
 F:388-415/Domain: DNA binding #stratus predicted <DNA>
 F:405-415/Region: helix-turn-helix

[illegible]

RESULT 6
G846534
Probable prolylcarboxypeptidase [imported] - *Arabisopsis thaliana*
C.Species: *Arabisopsis thaliana* (mouse-ear cress)
C.Date: 02-Feb-2001 #sequence_rev1000 02-Feb-2001 #next_change 16-Feb-2001
C.Accession: G846534
R.Lin, X.; Kahl, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, T.; Tallon, J.
euss, D.; Niemann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A.Title: Sequence and analysis of chromosome 2 of the plant *Arabisopsis thaliana*.
A.Reference number: A84420; M01D:20083487; PMID:1061197
A.Accession: G846534
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-476 <STO>
A.Cross-references: GB:AE002093; NID:g65598559; PIDN:AA18628.1; GSPDB:GN00139
C.Genetics:
A.Gene: At2g24280
A.Map position: 2
C.Superfamily: *Caenorhabditis elegans* ZK688.6 protein

Query Match	Similarity	38.8%	Score 47;	DB 2;	Length 476;
Best local	Similarity	42.9%	Pred. No. 20;		
Matches	9;	Conservative	5;	Mismatches	7;
				Indels	0;
				Gaps	0;
Qy	1 FPHHIFRGIVHVKTIHKLVT	21			
Db	410 YFHQIFRYLKNISSIVALT	430			

RESULT 7
AC1208
GTP-binding elongation factor homolog lmo1067 [imported] - *Listeria monocytogenes* (strain
C)Species: *Listeria monocytogenes*
C.Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #ext_change 14-Dec-2001
C.Accession: AC1208
R.Glasser, P.; Frangouli, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussuguet, O.; Entian, K.D.; Fshni, H.
D.; Jones, L.M.; Karszt, U.
Science 294, 849-852, 2001
A.Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Matsuura, E.; Maitournam, A.; Mat
ok, C.; Schuierer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.Title: Comparative genomics of *Listeria* species.
A.Reference number: AB1077, MUID:21537279, PMID:11679669
A.Accession: AC1208


```
C:Accession: T36329
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A:Reference number: 221575
A:Accession: T36329
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-201 <OUT>
A:Cross-references: EMBL:AL049841, PIDN:CA842756.1, GSPDB:GN00070, SCOEDB:SCE9.10C
A:Experimental source: Strain A3(2)
C:Genetics:
A:Gene: folE; SCOEDB:SCE9.10C
C:Superfamily: GTP cyclohydrolase I

Query Match      37.2%; Score 45; DB 2; Length 201;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY      3 HHI--FRGIHVHG 13
        |||: |||: |||
Db      92 HHLPFRGVAVHG 104

RESULT 13
E95088
dihydroorotate dehydrogenase A [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 17-May-2002
C:Accession: E95088
R:Telletlin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21557209; PMID:11463916
A:Accession: E95088
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74902.1; PID:g14972239; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0764
C:Superfamily: dihydroorotate oxidase

Query Match      36.4%; Score 44; DB 2; Length 311;
Best Local Similarity 52.6%; Pred. No. 36;
Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY      2 FHHIFRG--IVHVGKTIHK 18
        |||: |||: |||: |||
Db      257 FEHILGASWVQVGTTLHK 275

RESULT 14
H97955
dihydroorotate oxidase (EC 1.3.3.1) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: H97955
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain Re.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H97955
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99476.1; PID:g15458260; GSPDB:GN00174
```

```
C:Genetics:
A:Gene: pyrDA
C:Keywords: oxidoreductase

Query Match      36.4%; Score 44; DB 2; Length 311;
Best Local Similarity 52.6%; Pred. No. 36;
Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY      2 FHHIFRG--IVHVGKTIHK 18
        |||: |||: |||: |||
Db      257 FEHILGASWVQVGTTLHK 275

RESULT 15
C75553
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C:Accession: C75553
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75553
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <WHI>
A:Cross-references: GB:AE001878; GB:AE000513; NID:g6457820; PIDN:AAF09750.1; PID:g645782;
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0157
A:Map position: 1
C:Superfamily: Deinococcus radiodurans hypothetical protein DR0157

Query Match      36.4%; Score 44; DB 2; Length 513;
Best Local Similarity 45.8%; Pred. No. 64;
Matches 11; Conservative 1; Mismatches 4; Indels 8; Gaps 1;

QY      2 FHHIFRGIVH-----GKTIH 17
        |||: |||: |||: |||
Db      103 FHLPFGVVHVAHLPGRKDIH 126

Search completed: May 23, 2003, 08:57:18
Job time : 15 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:52:09 ; Search time 7.5 Seconds

(without alignments)
121.664 Million cell updates/sec

Title: US-09-929-788-3

Perfect score: 121
Sequence: 1 FFHHIRGIVHGKTIHLVTG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	39.7	351	1	RNPD_YERPE
2	46	38.0	501	1	NR52_CHICK
3	45	37.2	201	1	GCH1_STRCO
4	44	36.4	213	1	GCH1_OSTRO
5	44	36.4	311	1	PYRD_STRPN
6	44	36.4	905	1	RPA1_THRCE
7	43	35.5	192	1	EM3R_BACME
8	43	35.5	223	1	GCH1_CABEL
9	43	35.5	322	1	GDC_RAT
10	43	35.5	330	1	GDC_BOVIN
11	42.5	35.1	878	1	SV4_BUCAI
12	42	34.7	474	1	VSM5_TRYBB
13	42	34.7	865	1	RPA1_MERTW
14	42	34.7	873	1	RK_DROME
15	42	34.7	6486	1	TYCC_BACBR
16	41.5	34.3	235	1	ID11_CAMAC
17	41.5	34.3	533	1	NIFD_CLOPA
18	41.5	34.3	587	1	TALA_BFDV
19	41	33.9	124	1	NEUR_FELCA
20	41	33.9	236	1	GCH1_CHICK
21	41	33.9	241	1	GCH1_MOUSE
22	41	33.9	243	1	GCH1_YEAST
23	41	33.9	243	1	GCH1_HUMAN
24	41	33.9	250	1	GDC_HUMAN
25	41	33.9	332	1	DAD3_RHILQ
26	41	33.9	412	1	HUTH_THEAC
27	41	33.9	496	1	GSH1_BUCAI
28	41	33.9	518	1	PD14_MOUSE
29	41	33.9	666	1	VG1B_HSYND
30	41	33.9	865	1	RPA1_MERTW
31	41	33.9	870	1	VG1B_HSYND
32	41	33.9	917	1	ATB4_RAT
33	40.5	33.5	1203	1	ATB4_RAT

34	40	33.1	202	1	GCH1_MYCTU	O06273 mycobacteri
35	40	33.1	219	1	CLNT1_HUMAN	Q0bxu9 homo sapien
36	40	33.1	219	1	CLNT1_MOUSE	Q91jg7 mus musculu
37	40	33.1	222	1	RNC_ZYMO	Q925u2 zymomonas m
38	40	33.1	257	1	NUDC_ECO57	Q8x6x7 escherichia
39	40	33.1	257	1	NUDC_ECOLI	P32664 escherichia
40	40	33.1	257	1	NUDC_SALTI	O82328 salmonella
41	40	33.1	257	1	NUDC_SALTY	Q91915 salmonella
42	40	33.1	288	1	CD80_HUMAN	P33681 homo sapien
43	40	33.1	291	1	UPK_STAM	Q99vt8 staphylococ
44	40	33.1	291	1	UPK_STAV	Q9kins staphylococ
45	40	33.1	342	1	TRM1_THEVO	Q97ar2 thermoplasm

ALIGNMENTS

```

RESULT 1
RNPD_YERPE
ID RNPD_YERPE STANDARD; PRT; 351 AA.
AC Q8ZED2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfd.
GN RNPD OR YP02242.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tishball R.W., Holden M.T.G.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarrega A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,
CC rnfd, rnfe and rnfg (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE NORB/RNPD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: A014151; CAC91048.1; -
DR InterPro: IPR004338; NOR2_Rnfd_Rnfe.
DR Pfam: PF03116; NOR2_Rnfd_Rnfe; 1.
KW Electron transport; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 21
FT TRANSMEM 43
FT TRANSMEM 93
FT TRANSMEM 115
FT TRANSMEM 122
FT TRANSMEM 144
FT TRANSMEM 210
FT TRANSMEM 232
FT TRANSMEM 237
FT TRANSMEM 259
FT TRANSMEM 269
FT TRANSMEM 291
FT TRANSMEM 315
SQ SEQUENCE 351 AA; 38112 MW; A35EAE7524F605F8 CRC64;
Query Match 39.7%; Score 48; DB 1; Length 351;

```

Best Local Similarity 50.0%; Pred. No. 3.6;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 5 IFRGIVHVKTIHKLVYG 22
DB 160 IFTGYTHSGENHQLQVQ 177

RESULT 2

NR52_CHICK
ID NR52_CHICK STANDARD; PRT; 501 AA.
AC 042101;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Orphan nuclear receptor NR5A2 (OR2.0) (FTF/LRH-1).
GN NR5A2.
OS Gallus gallus (chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN NCBI [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97473520; PubMed=9332374;
RA Kudo T., Sutoh S.;
RT "Molecular cloning of chicken FTF-FI-related orphan receptors."
RL Gene 197261-266(1997).
CC -1- SUBCELLULAR LOCATION: Nuclear (probable).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NRS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB002403; BAA22838.1; -.
DR HSSP; P19793; ZML.
DR TRANSFAC; T04755; -.
DR InterPro; IPR000536; Hormone_rec_1lg.
DR InterPro; IPR001723; Strdhmn_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; ZNF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
FT DNA BIND 46 111 NUCLEAR RECEPTOR-TYPE.
FT ZN BIND 46 66
FT ZN BIND 82 106 C4-TYPE.
FT ZN BIND 115 144 C4-TYPE.
FT DOMAIN 115 144 FTF-FI BOX.
SQ SEQUENCE 501 AA; 57102 MW; 734596FC25682771 CRC64;

Query Match 38.0%; Score 46; DB 1; Length 501;
Best Local Similarity 52.6%; Pred. No. 11;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 HIFRGIVHVKTIHKLVYG 22
DB 350 HIYQVHVHKEGSIILVYG 368

RESULT 3

GCH1_STRCO

ID GCH1_STRCO STANDARD; PRT; 201 AA.

AC 09XR13;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).
GN FOLE OR SC03403 OR SCE9.10C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OC NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M45;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC (erythro)-1,2,3-trihydroxypropyl(dihydro)pyridine triphosphate.
CC -1- PATHWAY: Tetrahydrofolate biosynthesis; first step.
CC -1- SUBUNIT: HOMOPOLYMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL049841; CAB42756.1; -.
DR HSSP; P27511; IABR.
DR InterPro; IPR001474; GTP_cyclohydrol.
DR Pfam; PF01227; GTP_cyclohydrol; 1.
DR ProDom; PD003330; GTP_cyclohydrol; 1.
DR TIGRFAMs; TIGR00063; fole; 1.
DR PROSITE; PS00859; GTP_CYCLOHYDROL_1; 1.
DR PROSITE; PS00860; GTP_CYCLOHYDROL_1_2; 1.
KW One-carbon metabolism; Hydrolyase; Complete proteome.
FT DISULFID 90 163
SQ SEQUENCE 201 AA; 22415 MW; 4CC8FE6E76687B68 CRC64;

Query Match 37.2%; Score 45; DB 1; Length 201;
Best Local Similarity 61.5%; Pred. No. 6;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 3 HHI--FRGIVHVG 13
DB 92 HHLVPRGVAVHG 104

RESULT 4

GCH1_OSTOS
ID GCH1_OSTOS STANDARD; PRT; 213 AA.
AC 061573;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).
GN GCH.
OS Oostertagia ostertagi.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;

```

OC Trichostromyloidea; Haemochridae; Ostertagiinae; Ostertagia.
OX NCBI_TaxID=6317;
RN [1]
RP SEQUENCE FROM N.A.
RA Moore J., Devaney E.;
RT "Abundant messages from the 13 of Ostertagia ostertagi."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC (levythro-1,2,3-trihydroxypropyl)dihydropteridine triphosphate.
CC -1- SUBUNIT: Tetrahydrobiopterin biosynthesis, first step.
CC -1- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF052048; AAC06296.1; -.
DR HSSP: P27511; 1A8R.
DR InterPro: IPR001474; GTP_cyclohydrol.
DR Pfam: PF01227; GTP_cyclohydrol.1.
DR ProDom: PD003330; GTP_cyclohydrol.1.
DR TIGRFAMs: TIGR00063; FolE.1.
DR PROSITE: PS00859; GTP_CYCLOHYDROL_1_1.1.
DR PROSITE: PS00860; GTP_CYCLOHYDROL_1_2.1.
DR Tetrahydrobiopterin biosynthesis; Hydroxylase; Allosteric enzyme.
KM SEQUENCE 213 AA; 23843 MW; 2B5170D1902C32E9 CRC64;
SQ
Query Match 36.4%; Score 44; DB 1; Length 213;
Best Local Similarity 45.5%; Pred. No. 9.1;
Matches 10; Conservative 3; Mismatches 7; Indels 2; Gaps 1;
OY 3 HHT--PRGIVHGKTIHKLVGTG 22
Db 106 HHLVFPNGKVIHGYPNKKVLG 127

```

```

RESULT 5
PYRD STRPN STANDARD; PRT; 311 AA.
AC O9X50;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
DE (DHODHase) (DHODase) (DHOD).
GN PYRD OR PYRDA OR SP0764.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=1235/89;
RX MEDLINE=99362922; PubMed=10432287;
RA Lhull D., Munoz R., Lopez R., Garcia E.;
RT "A single gene (tss) located outside the cap locus directs the
RT formation of Streptococcus pneumoniae type 37 capsular polysaccharide.
RT Type 37 pneumococci are natural, genetically binary strains."
RL J. Exp. Med. 190:241-251(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Harte D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayem L.A., White O., Salzberg S.L., Lewis M.R., Rauden D.,
RA Holtapple E., Knouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,

```

```

RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC H(2)O(2).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- PATHWAY: Pyrimidine biosynthesis, fourth step.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.
CC SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ131985; CAB51330.1; -.
DR EMBL: AR007384; AAK74902.1; -.
DR HSSP: P54321; 2DOR.
DR TIGR: SP0764; -.
DR InterPro: IPR001295; DHO dh.
DR InterPro: IPR003009; FMN_enzyme.
DR Pfam: PF01180; DHODHase.1.
DR TIGRFAMs: TIGR01037; PYRD sub1_fam; 1.
DR PROSITE: PS00911; DHODHASE_1_1.
DR PROSITE: PS00912; DHODHASE_2_1.
DR Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KM Complete proteome.
FT NP BIND 242 250 FMN (POTENTIAL).
FT CONFLICT 174 174 H -> Y (IN RRF.1).
FT SEQUENCE 311 AA; 34527 MW; B01594DB3BF3752F CRC64;
SQ
Query Match 36.4%; Score 44; DB 1; Length 311;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 1;
OY 2 FHHIFRG--IYHVGKTIHK 18
Db 257 FEHLICGSMVQVGTITHK 275

```

```

RESULT 6
RPAL THECE STANDARD; PRT; 905 AA.
AC P31813;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6).
GN RPOA1.
OS Thermococcus celer.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=2264;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2476;
RX MEDLINE=93027167; PubMed=1408768;
RA Klenk H.-P., Schwass V., Lottepeich F., Zillig W.;
RT "Nucleotide sequence of the genes encoding the three largest subunits
RT of the DNA-dependent RNA polymerase from the archaeum Thermococcus
RL celer."
RL Nucleic Acids Res. 20:4659-4659(1992).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.

```

```

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- COFACTOR: ZINC.
CC -1- SIMILARITY: THE COMBINED 'A'+A" SUBUNITS CORRESPOND TO THE A
CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
CC EUBACTERIAL BETA' SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67313; CAA47723.1; -.
DR PIR; S25564; S25564.
DR HSSP; O9KWU6; IHOM.
DR InterPro; IPR000722; RNA_pol_A.
DR Pfam; PF00623; RNA_pol_A; 1.
DR Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;
KW Zinc-finger.
KW ZN FING.
SQ SEQUENCE 905 AA; 103 C4-TYPE (BY SIMILARITY);
Query Match 36.4%; Score 44; DB 1; Length 905;
Best Local Similarity 43.5%; Pred. No. 40;
Matches 10; Conservative 6; Mismatches 3; Indels 4; Gaps 2;
QY 2 FHFI--FRGIVHG--KTIHKLV 20
Db 74 FGHVELARPVHVGFAKTIHRLV 96
RESULT 7
BM3R_BACME STANDARD; PRT; 192 AA.
ID BM3R_BACME
AC P43506;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transcriptional repressor Bm3R1.
DE BM3R1.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92184811; PubMed=1544926;
RA Shaw G.C., Fulco A.J.;
RT "Barbiturate-mediated regulation of expression of the cytochrome
RT P450BM-3 gene of Bacillus megaterium by Bm3R1 protein.";
RL J. Biol. Chem. 267:5515-5526(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89291834; PubMed=2544578;
RA Rueckinger R.T., Wen L.P., Fulco A.J.;
RT "Coding nucleotide, 5' regulatory, and deduced amino acid sequences
RT of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450
RT reductase from Bacillus megaterium.";
RL J. Biol. Chem. 264:10987-10995(1989).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=93155125; PubMed=8428974;
RA Shaw G.C., Fulco A.J.;
RT "Inhibition by barbiturates of the binding of Bm3R1 repressor to its
RT operator site on the barbiturate-inducible cytochrome P450BM-3 gene
RT of Bacillus megaterium.";
RL J. Biol. Chem. 268:2997-3004(1993).
CC -1- FUNCTION: NEGATIVELY CONTROLS THE EXPRESSION OF THE CYTOCHROME
CC P450BM-3 GENE AT THE TRANSCRIPTIONAL LEVEL.
CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S87512; AAB21757.1; -.
DR EMBL; J04832; AAB87601.1; -.
DR InterPro; IPR01647; HTH_Tetr.
DR Pfam; PF00440; tetr; 1.
DR PRINTS; PR00455; HTHTETR.
DR PROSITE; PS01081; HTH_TETR_FAMILY; 1.
DR Transcription regulation; Repressor; DNA-binding.
KW DNA_BIND.
FT DNA_BIND 28 47
SQ SEQUENCE 192 AA; 21886 MW; 766AC6DD34944748 CRC64;
Query Match 35.5%; Score 43; DB 1; Length 192;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 FHFI--FRGIV 10
Db 84 FHFI--FRGIV 92
RESULT 8
GCHI_CAEEL STANDARD; PRT; 223 AA.
ID GCHI_CAEEL
AC Q19980;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).
DE F3268.6.
GN Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurtry A.;
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC (erythro-1,2,3-trihydroxypropyl) dihydropteridine triphosphate.
CC -1- PATHWAY: Tetrahydrobiopterin biosynthesis; first step.
CC -1- SUBUNIT: HOMOPOLYMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72509; CAA96650.1; -.
DR HSSP; P27511; IABR.
DR WormPep; F3268.6; CE05795.
DR InterPro; IPR001474; GTP_cyclohydrol; 1.
DR Pfam; PF01227; GTP_cyclohydrol; 1.
DR ProDom; PD003330; GTP_cyclohydrol; 1.
DR TIGRFAMs; TIGR00063; FolE; 1.
DR PROSITE; PS00859; GTP_CYCLOHYDROL_1; 1.
DR PROSITE; PS00860; GTP_CYCLOHYDROL_2; 1.
KW Hypothetical protein; Tetrahydrobiopterin biosynthesis; Hydrolyase;
KW Allosteric enzyme.
SQ SEQUENCE 223 AA; 25132 MW; B6584D6CBA6FA003 CRC64;

```

Query Match 35.5%; Score 43; DB 1; Length 223;
 Best Local Similarity 45.5%; Pred. No. 14;
 Matches 10; Conservative 3; Mismatches 7; Indels 2; Gaps 1;
 QY 3 HHI--FRGIVHVKTIHKLVGTG 22
 DB 115 HHLVPMGKVKHIGYIPNKKVLG 136

RESULT 9

GDC_RAT STANDARD; PRT; 322 AA.
 ID_GDC_RAT P16261;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Grave's disease carrier protein (GDC) (Mitochondrial solute carrier protein homolog) (Fragment).
 GN SL2C5A16 OR GDA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thyroid;
 RX MEDLINE=90114217; PubMed=2575220;
 RA Zarrilli R., Oates E.L., McBride O.W., Lerman M.I., Chan J.Y.,
 RA Santisteban P., Ursini M.V., Nockins A.L., Kohn L.D.;
 RT "Sequence and chromosomal assignment of a novel cDNA identified by
 RT immunoscreening of a thyroid expression library: similarity to a
 RL Mol. Endocrinol. 3:1498-1508 (1989).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (potential).
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M32973; AAA41639.1;
 DR InterPro; IPR001993; Mitoch_carr.
 DR Pfam; PF00153; mito_carr; 3.
 DR PROSITE; PS00215; MITOCH_CARRIER; 3.
 KW Mitochondrion; Transmembrane; Transport; Repeat.
 FT REPEAT 1 123 1.
 FT REPEAT 124 222 2.
 FT REPEAT 223 >322 3.
 FT NON_TER 322 322
 SQ SEQUENCE 322 AA; 35056 MW; F78CDA36CA9DC9A CRC64;

Query Match 35.5%; Score 43; DB 1; Length 322;
 Best Local Similarity 50.0%; Pred. No. 20;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 HIFRGIVHVKTIH 17
 DB 169 HTYGIHAFKITY 182

RESULT 10
 GDC_BOVIN STANDARD; PRT; 330 AA.
 AC 001888;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Grave's disease carrier protein (GDC) (Mitochondrial solute carrier protein homolog) OR GDC.
 GN SL2C5A16 OR GDA OR GDC.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxId=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart.
 RX MEDLINE=93091248; PubMed=1457817;
 RA Fiermonte G., Runswick M.J., Walker J.E., Palmieri F.;
 RT "Sequence and pattern of expression of a bovine homologue of a human,
 RT mitochondrial transport protein associated with Grave's disease.";
 RL DNA Seq. 3:71-78 (1992).
 CC -1- FUNCTION: Required for the accumulation of coenzyme A in the
 CC mitochondrial matrix (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (by similarity).
 CC -1- TISSUE SPECIFICITY: MOSTLY IN THYROID, LIVER, LUNG, KIDNEY AND
 CC TO A LESSER EXTENT IN HEART AND SKELETAL MUSCLE.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X66035; CAA46834.1;
 DR PIR; S26596; S26596.
 DR InterPro; IPR002067; Mit carrier.
 DR InterPro; IPR001993; Mitoch_carr.
 DR Pfam; PF00153; mito_carr; 3.
 DR PROSITE; PS00215; MITOCH_CARRIER; 3.
 KW Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
 FT TRANSMEM 132 152
 FT TRANSMEM 189 209
 FT TRANSMEM 232 252
 FT REPEAT 1 121 1.
 FT REPEAT 122 217 2.
 FT REPEAT 218 330 3.
 SQ SEQUENCE 330 AA; 36085 MW; 4C614701D2B8DEA9 CRC64;

Query Match 35.5%; Score 43; DB 1; Length 330;
 Best Local Similarity 50.0%; Pred. No. 20;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 HIFRGIVHVKTIH 17
 DB 164 HTYGIHAFKITY 177

RESULT 11
 SYA_BUCAL STANDARD; PRT; 878 AA.
 AC P57483;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
 GN ALAS OR BUA03.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxId=118099;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC SRRAIN-Tokyo 1998;
RX MEDLINE=20455133; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. Abs.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001119; BAB13106.1;
DR InterPro; IPR002106; AAcRNA_ligaseII.
DR InterPro; IPR003156; DHAL.
DR InterPro; IPR002318; tRNA-synt 2c.
DR Pfam; PF01411; tRNA-synt 2c; 1.
DR Pfam; PF02272; DHAL; 1.
DR Pfam; PR00980; TRNASYTTHALA.
DR PRINTS; TIGR00344; alas; 1.
DR TIGRPFAM; TIGR00344; alas; 1.
DR PROSITE; PS50860; AA tRNA_LIGASE_II Ala; 1.
KW Aminoacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding;
KW complete proteome.
SQ SEQUENCE 878 AA; 101402 MW; 005321303125D165 CRC64;

Query Match 35.1%; Score 42.5; DB 1; Length 878;
Best Local Similarity 50.0%; Pred. No. 66; 4; Indels 1; Gaps 1;
Matches 8; Conservative 3; Mismatches 4;

QY 1 FPHIFRGIVHV-GKT 15
Db 330 FPHKLVSVIHWGKT 345

RESULT 12
VSM5 TRYBB STANDARD; PRT; 474 AA.
ID _VSM5_TRYBB
AC P26333;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Variant surface glycoprotein MITR1.5 precursor (VSG 118).
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Isolate MITG 118;
RX MEDLINE=92046037; PubMed=1942032;
RA Carrington M., Miller N., Blum W.L., Roditi I., Wiley D.C.;
RA Turner M.J.;
RT "Variant specific glycoprotein of Trypanosoma brucei consists of two
RT domains each having an independently conserved pattern of cysteine
RT residues.";
RL J. Mol. Biol. 221:823-835(1991).
CC -1- FUNCTION: VSG FORMS A COAT ON THE SURFACE OF THE PARASITE. THE
CC TYRANOSOME EVADES THE IMMUNE RESPONSE OF THE HOST BY EXPRESSING
CC A SERIES OF ANTIGENICALLY DISTINCT VSgs FROM AN ESTIMATED 1000
CC VSG GENES.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC A SOLUBLE FORM IS RELEASED FROM RUPTURED CELLS BY THE ACTION OF A
CC PI-PLC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56763; CAA40082.1; -.
DR PIR; S18452; S18452.
DR InterPro; IPR001812; Trypan_glycop.
DR Pfam; PF00913; Trypan_glycop; 1.
KW Glycoprotein; Antigen; Trypanosomiasis; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 451
FT PROPEP 452 474
FT DISULFID 37 161
FT DISULFID 144 214
FT CARBOHYD 74 74
FT CARBOHYD 95 95
FT CARBOHYD 329 329
FT CARBOHYD 451 451
FT LIPID 451 451
SQ SEQUENCE 474 AA; 49754 MW; 5DD75808A01BB6BB CRC64;

Query Match 34.7%; Score 42; DB 1; Length 474;
Best Local Similarity 38.9%; Pred. No. 42; 7; Indels 0; Gaps 0;
Matches 7; Conservative 4; Mismatches 3;

QY 4 HIFRGIVHVKTIHKLV 21
Db 325 HIFLVSHNLRELQAVT 342

RESULT 13
RPAL METTW STANDARD; PRT; 865 AA.
ID _RPAL_METTW
AC P09846;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6).
DN RPOA1 OR RPOAT.
GN Methanobacterium thermoautotrophicum (strain Winter).
OS Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=79930;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88335550; PubMed=2843811;
RX Bergroeff B., Koeckel L., Koertner C., Truss M., Schallenberg J.,
RA Klein A.;
RT "Relatedness of archaeobacterial RNA polymerase core subunits to their
RT eubacterial and eukaryotic equivalents.";
RL Nucleic Acids Res. 16:8113-8128(1988).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -1- COFACTOR: ZINC.
CC -1- SUBUNIT: M.THERMOAUTOTROPHICUM RNAP IS COMPOSED OF FOUR SUBUNITS:
CC A', A'', B' AND B''.
CC -1- SIMILARITY: THE COMBINED A'+A'' SUBUNITS CORRESPOND TO THE A
CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
CC EUBACTERIAL BETA' SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X08038; CAA30838.1; -.

```



```

DR PIR, S02196; S02196.
DR InterPro; IPR000722; RNA_pol_A.
DR Pfam; PF00623; RNA_pol_A; 1.
DR KMW; Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW Zinc-finger.
FT ZN FING 60 103 C4-TYPE (POTENTIAL).
SQ SEQUENCE 865 AA; 97518 MW; 33C65710C6A4E7D6 CRC64;

Query Match 34.7%; Score 42; DB 1; Length 865;
Best Local Similarity 40.0%; Pred. No. 78;
Matches 8; Conservative 7; Mismatches 3; Indels 2; Gaps 1;

Oy 4 HIFRQIVHG-KITHKLVY 21
Db 78 NLARPVHVGFDTHKILS 97

RESULT 14
RX DROME STANDARD; PRT; 873 AA.
AC Q9W2Q1; Q46035;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinal homeobox protein Rk (DRX1) (DRX).
GN RX OR CG10052.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;

[1]
RX SEQUENCE FROM N.A.
RC STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=98151514; PubMed=9482887;
RA Eggert T., Hauck B., Hildebrandt N., Gehring W.J., Waldorf U.;
RT "Isolation of a Drosophila homolog of the vertebrate homeobox gene Rx
and its possible role in brain and eye development.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:2343-2348(1998).

[2]
RX SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-U., Andrews-Plankkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cwley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Uralai M., Kalush F., Kapen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Pacleby J.M.,
RA Palazzolo M., Peltan G.S., Pan S., Nusskern D.R., Pacleby J.M.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kamos I., Simpson M.D., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

```

RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye U., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

[3]
RN
RP REVISIONS.
RC STRAIN=Berkley;
RA Misra S., Crosby M.A., Mathews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celisner S.E.,
RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.U.,
RA Harris N.L., Krommiller B., Marshall B., Millburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Shu S., Smutnick F.,
RA Whitfield E.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.

[4]
RN PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=97320497; PubMed=9177348;
RA Mathers P.H., Gribner A., Mahon K.A., Jamrich M.;
RT "The Rx homeobox gene is essential for vertebrate eye development.";
RL Nature 387:603-607(1997).

CC - FUNCTION: Appears to function in brain development.
CC - SUBCELLULAR LOCATION: Nuclear (By similarity).
CC - DEVELOPMENTAL STAGE: Expressed in the procephalic region and in
CC the clypeolabrum from stage 8 on and later in the brain and the
CC central nervous system.
CC - SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC - "BICOID" SUBFAMILY.
CC - SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC - CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL; AJ223300; CA11241.1; ALT INIT.
DR EMBL; AE003452; AAF4639.2; ALT_SEQ.
DR TRANSFAC; T03511; -.
DR FLYBASE; FBgn0020617; Rx.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 116 123 OCTAPEPTIDE MOTIF.
FT DNA_BIND 526 586 HOMEBOX.
FT FT 849 862 OAR.
FT DOMAIN 855 859 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 329 333 POLY-PRO.
FT DOMAIN 366 369 POLY-ASP.
FT DOMAIN 437 453 POLY-GLN.
FT DOMAIN 670 673 POLY-PRO.
FT DOMAIN 683 691 POLY-PRO.
FT DOMAIN 753 763 POLY-SER.
FT DOMAIN 828 834 POLY-PRO.
FT CONFLICT 4 4 S -> P (IN REF. 1).

```

```

FT CONFLICT 136 136 R -> W (IN REF. 1).
FT CONFLICT 384 384 S -> T (IN REF. 1).
FT CONFLICT 645 672 PLSIAPGNLMTSLAMGHMHANNGPPP -> QPGARKSDH
FT CONFLICT 768 768 EDSGCHGPPPCPOMAAA (IN REF. 1).
SQ SEQUENCE 873 AA; 92897 MW; 87B56AD693F610 CRC64;

Query Match 34.7%; Score 42; DB 1; Length 873;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 FHIIFRGIVHVGKTIHKLT 19
DB 23 FQHIIEQVLVQGGGNHKL 40

RESULT 15
TYCC_BACBR STANDARD; PRT; 6486 AA.
ID TYCC_BACBR STANDARD; PRT; 6486 AA.
AC 030409;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase III [Includes: ATP-dependent asparagine
DE adenylyase (AsnA) (Asparagine activase); ATP-dependent glutamine
DE adenylyase (GlnA) (Glutamine activase); ATP-dependent tyrosine
DE adenylyase (TyrA) (Tyrosine activase); ATP-dependent valine adenylyase
DE (ValA) (Valine activase); ATP-dependent ornithine adenylyase (OrnA)
DE (Ornithine activase); ATP-dependent leucine adenylyase (LeuA) (Leucine
DE activase)].
GN TYCC.
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 8185;
RC MEDLINE=98012987; PubMed=9352938;
RX Mootz H.D., Marshiel W.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
RT nucleotide sequence and biochemical characterization of functional
RT internal adenylation domains.";
RL J. Bacteriol. 179:6843-6850 (1997).
CC -! FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
CC GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
CC PEPTIDE PRODUCT.
CC -! COFACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTETHEINES (BY
CC SIMILARITY).
CC -! PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -! SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF TYCA, TYCB AND TYCC.
CC -! DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC ADENYLYATION, THIOYLATION, CONDENSATION (NOT FOR THE INITIATION
CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC (OPTIONAL).
CC -! MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -! SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -! SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; AF004835; AAC45930.1; -.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantne attach.
DR InterPro; IPR000379; Ser ests site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 6.
DR Pfam; PF00550; pp-binding; 6.
DR Pfam; PF00668; Condensation; 6.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMBIBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR PROSITE; PS00455; AMP-BINDING; 6.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
KW Ligase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.
FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
FT BINDING 1000 1000 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2037 2037 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3075 3075 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4110 4110 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 5154 5154 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 6197 6197 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 6486 AA; 724011 MW; 4934900AF07DE786 CRC64;

Query Match 34.7%; Score 42; DB 1; Length 6486;
Best Local Similarity 45.0%; Pred. No. 6.1e+02;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 FHIIFRGIVHVGKTIHKLT 21
DB 141 FHIILVQVLTGLVHLKLLT 160

```

Search completed: May 23, 2003, 08:55:47
Job time : 9.5 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:52:40 ; Search time 24.5 Seconds

(without alignments)
185.022 Million cell updates/sec

Title: US-09-929-788-3

Perfect score: 121

Sequence: 1 FFHIFRGIVHWKTIHKLVTC 22

Scoring table: BLOSUM62

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SPTREMBL.21.*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_uniclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	121	100.0	79	13	Q8UUG2	Q8UUG2 morone chry
2	118	97.5	79	13	Q8UUG0	Q8UUG0 morone saxa
3	51	42.1	436	2	Q45144	Q45144 corynebacte
4	49.5	40.9	698	2	Q92486	Q92486 aeromonas p
5	48	39.7	351	16	Q8ZED2	Q8ZED2 yersinia pe
6	48	39.7	404	3	Q12090	Q12090 saccharomyc
7	47	38.8	28	2	Q9RE14	Q9RE14 acidiphiliu
8	47	38.8	173	17	Q29424	Q29424 archaeosporiob
9	47	38.8	436	2	Q45293	Q45293 corynebacte
10	47	38.8	476	10	Q8S810	Q8S810 arabidopsis
11	47	38.8	612	16	Q92C66	Q92C66 listeria in
12	47	38.8	612	16	Q8Y851	Q8Y851 listeria mo
13	46.5	38.4	679	12	Q991E8	Q991E8 avian nephri
14	46	38.0	207	5	Q8WT62	Q8WT62 aedes aegypt
15	46	38.0	537	10	Q9FW78	Q9FW78 oryza sativ
16	46	38.0	907	17	O59227	O59227 pyrococcus

17	46	38.0	907	17	Q9V114	Q9V114 pyrococcus
18	45.5	37.6	174	10	Q9S0T8	Q9S0T8 arabidopsis
19	45.5	37.6	476	17	Q8THC2	Q8THC2 methanosarc
20	45.5	37.6	480	4	Q96E59	Q96E59 homo sapien
21	45.5	37.6	529	4	Q96PD9	Q96PD9 homo sapien
22	45	37.2	39	4	Q8WKK8	Q8WKK8 homo sapien
23	45	37.2	207	5	Q9GPE7	Q9GPE7 aedes aegypt
24	45	37.2	207	5	Q9GPE6	Q9GPE6 aedes aegypt
25	45	37.2	207	5	Q8WT61	Q8WT61 aedes aegypt
26	45	37.2	207	5	Q8TSC5	Q8TSC5 aedes aegypt
27	45	37.2	245	16	Q8XSM3	Q8XSM3 raietonia s
28	45	37.2	269	4	Q9HA88	Q9HA88 homo sapien
29	45	37.2	502	4	Q9P233	Q9P233 homo sapien
30	45	37.2	816	16	Q98P52	Q98P52 rhizobium 1
31	45	37.2	1775	13	Q90XG6	Q90XG6 brachydanto
32	44.5	36.8	358	10	Q40877	Q40877 petunia hyb
33	44	36.4	75	2	Q8RO18	Q8RO18 bacillus ce
34	44	36.4	117	2	Q9LAN0	Q9LAN0 streptococ
35	44	36.4	182	13	Q90YGS	Q90YGS oncorhynch
36	44	36.4	202	5	Q9U8Y6	Q9U8Y6 tachypleus
37	44	36.4	209	17	Q8TQ49	Q8TQ49 methanosarc
38	44	36.4	372	12	Q9PYV2	Q9PYV2 xestia c-ni
39	44	36.4	456	10	Q8VYU3	Q8VYU3 lycopersico
40	44	36.4	478	10	Q9P6M6	Q9P6M6 lycopersico
41	44	36.4	480	5	Q9VJ45	Q9VJ45 drosophila
42	44	36.4	513	16	Q9RXZ7	Q9RXZ7 delnoscoccus
43	44	36.4	517	5	Q8T063	Q8T063 drosophila
44	44	36.4	859	15	Q73307	Q73307 human immun
45	44	36.4	952	16	Q9KUG7	Q9KUG7 vibrio chol

ALIGNMENTS

RESULT 1

Q8UUG2 PRELIMINARY; PRT; 79 AA.
AC Q8UUG2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Moronecidin prepeptide precursor.
OS Morone chrysopt (white bass).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes; Percoidae;
OC Moronidae; Morone.
OX NCBI_TaxID=46259;
RN [1]
RP SEQUENCE FROM N.A.
RA Lauch X.C., Shike H., Burns J.C., Westerman M.E., Ostland V.E.,
RA Carlberg J.M., Van Olst J.C., Nizet V., Taylor S.W., Shimizu C.,
RA Bulet P.,
RT "Discovery and characterization of two isoforms of moronecidin, a
RT novel antimicrobial peptide from hybrid striped bass."
RL J. Biol. Chem. 0:0-0(2002).
DR EMBL; AF394243; AAL57318.1; -.
KW SIGNAL.
FT SIGNAL. 1 22 POTENTIAL.
FT CHAIN 23 44 MORONECIDIN.
SQ SEQUENCE 79 AA; 9106 MW; B6F32481C4200EBB CRC64;
Query Match 100.0%; Score 121; DB 13; Length 79;
Best Local Similarity 100.0%; Pred. No. 6e-12;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FFHHIFRGIVHGKTIHKLVTG 22
 |||||
 Db 23 FFHHIFRGIVHGKTIHKLVTG 44

RESULT 2

Q8UUG0 PRELIMINARY; PRT; 79 AA.

AC Q8UUG0.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Moronecidin.
 OS Morone saxatilis (Striped bass).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
 OC Moronidae; Morone.
 OC NCB1_TaxID=34816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lachn X., Shike H., Burns J.C., Westerman M.E., Ostland V.E.,
 RA Carlborg J.M., Van Olst J.C., Nizet V., Taylor S.W., Shimizu C.,
 RA Bulet P.;
 RT "Discovery and characterization of two isoforms of moronecidin, a
 RT novel antimicrobial peptide from hybrid striped bass.";
 RL J. Biol. Chem. 0:0-0(2002).
 DR EMBL; AF394244; AAL57319.1; -
 DR EMBL; AF385583; AAL49496.1; -
 SQ SEQUENCE 79 AA; 9222 MW; C9DF13BD4FA7EFD CRC64;

Query Match 97.5%; Score 118; DB 13; Length 79;
 Best Local Similarity 95.5%; Pred. No. 1.8e-11;
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FFHHIFRGIVHGKTIHKLVTG 22
 |||||
 Db 23 FFHHIFRGIVHGKTIHKLVTG 44

RESULT 3

Q45144 PRELIMINARY; PRT; 436 AA.

AC Q45144.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE DNA, transposable element IS31831 (Transposase).
 GN TPN.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.
 OC NCB1_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC31831;
 RX MEDLINE=94254729; PubMed=8196545;
 RA Vertes A., Inui M., Kobayashi M., Kurusu Y., Yukawa H.;
 RT "Isolation and characterization of IS31831 a transposable element from
 RT Corynebacterium glutamicum.";
 RL Mol. Microbiol. 11:739-746(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CNE;
 RA Nihomiyu K., Garbe T.R., Inui M., Yukawa H.;
 RT "Survival of High-Dose Cyanide Associated with Transposition in
 RT Corynebacterium glutamicum.";
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D17429; BAA04250.1; -
 DR EMBL; AF435074; AAL31543.1; -

DR EMBL; AF435074; AAL31536.1; -
 DR InterPro; IPR002560; Transposase_12.
 DR Pfam; PF01610; Transposase_12; 1.
 KW DNA-binding.
 SQ SEQUENCE 436 AA; 49576 MW; 116D3A78DCADBSA CRC64;

Query Match 42.1%; Score 51; DB 2; Length 436;
 Best Local Similarity 41.2%; Pred. No. 7;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HHIFRGIVHGKTIHKL 19
 |||||
 Db 417 HYLRLHSGQLVHKL 433

RESULT 4

Q9Z486 PRELIMINARY; PRT; 698 AA.

AC Q9Z486;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Alpha-glucuronidase.
 GN XYG.
 OS Aeromonas punctata (Aeromonas caviae).
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OC NCB1_TaxID=648;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=ME-1;
 RC Kitagawa E., Suzuki T., Kawai K.;
 RT "Xylosidase gene(xysB) and alpha-glucuronidase gene(xyg) of Aeromonas
 RT caviae ME-1.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB022788; BAA74508.1; -
 DR InterPro; IPR005154; Glyco_hydro_67.
 DR Pfam; PF03648; Glyco_hydro_67; 1.
 SQ SEQUENCE 698 AA; 78166 MW; 064958F833654252 CRC64;

Query Match 40.9%; Score 49.5; DB 2; Length 698;
 Best Local Similarity 47.1%; Pred. No. 20;
 Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

Qy 1 FFHHIFRGIVHGKTI 16
 |||||
 Db 623 FFHHVPTVHLSGKTV 639

RESULT 5

Q8ZED2 PRELIMINARY; PRT; 351 AA.

AC Q8ZED2;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative integral membrane protein.
 GN YP02242.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OC NCB1_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holtroyd S., Jagsis K., Kariyeh A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RA Nature 413:523-527(2001).
 DR EMBL; AJ14151; CAC91048.1; -
 DR InterPro; IPR004338; NOR2_RnFD_RnFE.
 DR Pfam; PF03116; NOR2_RnFD_RnFE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 351 AA; 38112 MW; A35EAE7524F605F8 CRC64;

Query Match 39.7%; Score 48; DB 16; Length 351;
 Best Local Similarity 50.0%; Pred. No. 17;
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 5 IPRGVHVKTHKLVTVG 22
 DB 160 IFTGYTHSGENIHOLVVG 177

RESULT 6
 O12090 PRELIMINARY; PRT; 404 AA.
 ID O12090
 AC O1-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Similar to S. POMBE hypothetical protein C22G7.04P (YLR107WP).
 GN REX3 OR L8004.1 OR L2904 OR YLR107W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Geisel C.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier T., Johnson D.,
 RA Johnston L., Langston Y., Lacroille P., Le T., Mardis E., Menzies S.,
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
 RA Taich A., Trevaakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
 RA Wilson R., Waterston R.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Waterston R.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FY23 /RD005;
 RA Verhaesselt P., Volckaert G.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SEQUENCE OF 275-404 FROM N.A.
 RC STRAIN=S288C;
 RA MEDLINE=97313267; Pubmed=9169871;
 RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Anseorge W.,
 RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
 RA Eutian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Haus-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter P.,
 RA Louis E.J., Messenguy F., Mewes H.W., Miosga T., Mostl D.,
 RA Muller-Auer S., Nentwig U., Obermaier B., Piravandi E., Pohl T.M.,
 RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
 RA Scharte M., Scherens B., Scholler P., Schwarz S., Schwarz S.,
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhaesselt P.,

RA Viendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Heise J.D.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII."
 RL Nature 387:0-0(0).
 RN [7]
 RP SEQUENCE OF 275-404 FROM N.A.
 RC STRAIN=S288C;
 RA Cherry J.M.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53876; AAB67549.1; -
 DR EMBL; X89514; CAA61685.1; -
 DR EMBL; Z73279; CAA97672.1; -
 DR EMBL; U53878; AAB67561.1; -
 DR EMBL; Z73280; CAA97675.1; -
 DR SGD; S0004097; REX3.
 DR InterPro; IPR000520; Exonuclease.
 DR Pfam; PF00929; Exonuclease; 1.
 DR SMART; SM00479; EXOIII; 1.
 SQ SEQUENCE 404 AA; 45879 MW; FFEDF994ECA69A5F CRC64;

Query Match 39.7%; Score 48; DB 3; Length 404;
 Best Local Similarity 87.5%; Pred. No. 19;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FFHIFRG 8
 DB 212 FFHIFRG 219

RESULT 7
 O9RE14 PRELIMINARY; PRT; 28 AA.
 ID O9RE14
 AC O9RE14;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE Hypothetical 3.4 kDa protein.
 OS Acidiphilium symbioticum.
 OC Acidiphilium symbioticum.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Acidiphilium.
 OX NCBI_TaxID=94005;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KM2;
 RA Mahapatra N.R., Ghosh S., Deb C., Banerjee P.C.;
 RT "Cloning, sequencing and analysis of a Cadmium and zinc resistance
 RT conferring plasmid segment from Acidiphilium symbioticum KM2."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ239066; CAB65723.1; -
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 28 AA; 3379 MW; CA15E234000EC599 CRC64;

Query Match 38.8%; Score 47; DB 2; Length 28;
 Best Local Similarity 41.2%; Pred. No. 16;
 Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 3 HHIFRGVHVKTHKLV 19
 DB 6 HRLDQIRHGLGTFHRL 22

RESULT 8
 O29424 PRELIMINARY; PRT; 173 AA.
 ID O29424
 AC O29424;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Ferritin, putative.
 GN AF0834.
 OS Archaeoglobus fulgidus.
 OC Archaeae; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OC Archaeoglobaceae; Archaeoglobus.
 OK NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=8049343; Pubmed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyriades N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 RU EMBL; AB001047; AAB90406.1; -.
 DR HSSP; P2387; IEUM.
 DR TIGR; AF0834; -.
 DR InterPro; IPR001519; Ferritin.
 DR Pfam; PF00210; ferritin; 1.
 KM Hypothetical protein: Complete proteome.
 SQ SEQUENCE 173 AA; 20316 MW; DD555B59B7A30F5 CRC64;

Query Match 38.8%; Score 47; DB 17; Length 173;
 Best Local Similarity 47.4%; Pred. No. 11;
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

RESULT 9
 045293 PRELIMINARY; PRT; 436 AA.
 ID 045293;
 AC 045293;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Transposase.
 OS *Corynebacterium glutamicum* (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 OC *Corynebacterium*.
 OK NCBI_TaxID=1718;
 GN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13869;
 RX MEDLINE=96200862; Pubmed=8621097;
 RA Correia A., Pisabarro A., Castro J., Martin J.;
 RT "Cloning and characterization of an IS-like element present in the
 RT genome of *Brevibacterium lactofermentum* ATCC 13869.";
 RL Gene 170:91-94(1996).
 DR EMBL; Z66534; CA931438.1; -.
 DR InterPro; IPR002560; Transposase_12.
 DR Pfam; PF01610; Transposase_12; 1.
 SQ SEQUENCE 436 AA; 49406 MW; 2568EC0A94A1E773 CRC64;

Query Match 38.8%; Score 47; DB 2; Length 436;
 Best Local Similarity 41.2%; Pred. No. 31;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 HHIFRGIVHVKTIHKLV 19
 Db 417 HYILRLIHSGQLTHKI 433

RESULT 10

08S8L0
 ID 08S8L0 PRELIMINARY; PRT; 476 AA.
 AC 08S8L0;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative polycarboxypeptidase.
 GN ATG24280.
 OS *Arabidopsis thaliana* (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005967; AAM15096.1; -.
 KW Carboxypeptidase.
 SQ SEQUENCE 476 AA; 54423 MW; 3C3FE8E1F8A580D2 CRC64;

Query Match 38.8%; Score 47; DB 10; Length 476;
 Best Local Similarity 42.9%; Pred. No. 34;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

RESULT 11
 092CW6 PRELIMINARY; PRT; 612 AA.
 ID 092CW6;
 AC 092CW6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein lin1055.
 GN LIN1055.
 OS *Listeria innocua*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OK NCBI_TaxID=1642;
 GN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / SEROVAR 6A;
 RX Pubmed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Chardib A., Chetouani F., Couve E., de Darvar A., Denoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entlian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkak G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
 RT "Comparative genomics of *Listeria species*.";
 RL Science 294:849-852(2001).
 DR EMBL; AL596167; CAC96286.1; -.


```

DT 01-JUN-2002 (Tremblrel 21, last annotation update)
DE Putative cytochrome P450.
GN OSJNBA0026L12.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OK NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPONBARE;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Hsiao J., Ziemann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Utechtack T.R., Khatak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBA0026L12 genomic sequence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AC068924; AAC13500.1; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KM Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 537 AA; 59970 MW; 760CF5A958C27383 CRC64;

Query Match 38.0%; Score 46; DB 10; Length 537;
Best Local Similarity 46.7%; Pred. No. 56;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 HIFRGIVHVGKTIHK 18
   |||||
Db 238 HLFGRVH/GLAVRR 252

```

Search completed: May 23, 2003, 08:56:44
 Job time : 25.5 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:50:29 ; Search time 31 Seconds
(without alignments)
94.565 Million cell updates/sec

Title: US-09-929-788-2
Perfect score: 121
Sequence: 1 FFHHIFRGIVHGKTHRLVTG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	22	23	AAE20173
2	121	100.0	22	23	ABR07721
3	118	97.5	22	23	AAE20174
4	118	97.5	22	23	ABR07722
5	87	71.9	22	23	AAE20172
6	87	71.9	22	23	ABR07720
7	65	53.7	22	23	AAE20177
8	49	40.5	907	22	AAE20197
9	48.5	40.1	290	21	AAE20499
10	48.5	40.1	392	22	AAE20392

11	48.5	40.1	392	22	AAE20393	Stem cell growth f
12	48.5	40.1	425	22	AAE20394	Human polypeptide
13	48.5	40.1	425	22	AAE20395	Stem cell growth f
14	48.5	40.1	427	22	AAE20396	Human polypeptide
15	48.5	40.1	427	22	AAE20397	Human polypeptide
16	48.5	40.1	449	22	AAE20398	Stem cell growth f
17	48.5	40.1	486	22	AAE20399	Human polypeptide
18	48.5	40.1	529	22	AAE20400	Stem cell growth f
19	48.5	40.1	529	22	AAE20401	Human PEO polypept
20	48.5	40.1	529	22	AAE20402	Human polypeptide
21	48.5	40.1	529	22	AAE20403	Stem cell growth f
22	48.5	40.1	529	22	AAE20404	Amino acid sequenc
23	48.5	40.1	529	22	AAE20405	Human Tumour Endoc
24	48.5	40.1	529	22	AAE20406	Human Tumour Endoc
25	48.5	40.1	529	22	AAE20407	Human Immune/Naema
26	48.5	40.1	529	22	AAE20408	Protonibacterium
27	48.5	40.1	529	22	AAE20409	ISG1 ORF-2 prod.
28	48.5	40.1	529	22	AAE20410	Corynebacterium gl
29	48.5	40.1	529	22	AAE20411	Corynebacterium gl
30	48.5	40.1	529	22	AAE20412	Protein product of
31	48.5	40.1	529	22	AAE20413	C glutamicum prote
32	48.5	40.1	529	22	AAE20414	C glutamicum prote
33	48.5	40.1	529	22	AAE20415	C glutamicum prote
34	48.5	40.1	529	22	AAE20416	Listeria monocytrog
35	48.5	40.1	529	22	AAE20417	C glutamicum prote
36	48.5	40.1	529	22	AAE20418	Corynebacterium gl
37	48.5	40.1	529	22	AAE20419	Corynebacterium gl
38	48.5	40.1	529	22	AAE20420	Arabidopsis thalia
39	48.5	40.1	529	22	AAE20421	Arabidopsis thalia
40	48.5	40.1	529	22	AAE20422	Arabidopsis thalia
41	48.5	40.1	529	22	AAE20423	Novel ORFX protei
42	48.5	40.1	529	22	AAE20424	Novel human diagno
43	48.5	40.1	529	22	AAE20425	Human protein sequ
44	48.5	40.1	529	22	AAE20426	C glutamicum prote
45	48.5	40.1	529	22	AAE20427	Corynebacterium gl

ALIGNMENTS

RESULT 1
AAE20173
ID AAE20173 standard; peptide; 22 AA.
XX AAE20173;
AC
XX 18-JUN-2002 (first entry)
DT
XX Fish mast cell antimicrobial peptide, piscidin 1 (P1).
DE
XX Antimicrobial peptide; endobiotic peptide; mast cell; piscidin; fish;
XX methicillin; vancomycin; streptogramin; microbial infection; stress;
KW bacterial disease; fish health; screening; gene therapy; antimicrobial;
KW tranquilliser; fungicide; protozoacide; antibacterial; antiparasitic.
XX
XX Morone chrysops.
OS Morone saxatilis.
XX
XX WO200214345-A2.
PN
XX 21-FEB-2002.
PD
XX 13-AUG-2001; 2001WO-US41696.
PF
XX 15-AUG-2000; 2000US-225354P.
PR
XX (UYNC-) UNIV NORTH CAROLINA STATE.
PA
XX Noga EJ, Silphaduang U;
PI Antimicrobial pep
XX WPI; 2002-269176/31.
DR
XX Novel antimicrobial peptides (endobiotic peptides) isolated from mast
PT

PT cells, specifically fish mast cells, useful for treating microbial
PT infections in humans or animals and for reducing antibiotic resistance
XX in bacteria -
XX
XX Claim 4; Page 17; 31pp; English.
XX
CC The invention relates to antimicrobial peptide (endobiotic peptide)
CC isolated from a mast cell. Such peptides obtained from fish mast cells
CC are referred as piscidins. The antimicrobial peptide is useful for
CC treating microbial infection in a subject, and for reducing antibiotic
CC (e.g., methicillin, vancomycin or streptogramin) resistance in bacteria
CC such as *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus aureus*,
CC *Klebsiella pneumoniae*, *Streptococcus faecalis* or *Shigella flexneri*. These
CC peptides are useful for treating microbial infections in human or animals
CC and for preparing a medicament for treating microbial infections. These
CC peptides are also useful for treating infections caused by microorganisms
CC other than bacteria, e.g., fungi, yeast, protozoa, parasite, etc. These
CC peptides are useful for treating stress or bacterial disease in fish and
CC monitoring fish health. The inverse relationship between endobiotic
CC levels and stress also allows assessment of freshness of fish food
CC product. Nucleic acid sequences encoding these peptides are useful for
CC diagnostic purposes; treating stress in a fish and for screening
CC additional endobiotic peptides. They are also useful in gene therapy.
CC The present sequence is piscidin 1 (P1) peptide isolated from the gills
CC of hybrid striped bass (Morone saxatilis x Morone chrysops).
XX
SQ Sequence 22 AA;
XX
Query Match 100.0%; Score 121; DB 23; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.5e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FFHHIFRGIVHWGKTIHRLVTG 22
DB 1 FFHHIFRGIVHWGKTIHRLVTG 22
XX
RESULT 2
ABBO7721
ID ABB07721 standard; peptide; 22 AA.
XX
AC ABB07721;
XX
DT 10-JUN-2002 (first entry)
XX
DE Antimicrobial peptide #2 from hybrid striped bass.
XX
KM Fish; hybrid striped bass; endobiotic; antibiotic; tranquilizer;
XX antimicrobial; antibacterial; fish food product.
XX
OS Morone saxatilis x Morone chrysops.
XX
PN WO200214346-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US41697.
XX
PR 15-AUG-2000; 2000US-225354P.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Noga EJ, Silphaduang U;
XX
DR WPI; 2002-269177/31.
XX
PT Novel antimicrobial compound (endobiotic peptide isolated from fish),
PT useful for monitoring fish health and monitoring freshness of fish food
PT product and for screening compounds which treat stress in fish -
XX
PS Claim 1; Page 16; 25pp; English.
CC The invention provides antimicrobial peptides (ABB07720-23) isolated from

CC hybrid striped bass. The peptides were purified from the endobiotic
CC family. The endogenous endobiotic peptides exhibiting antimicrobial
CC activity) are useful for monitoring fish health, monitoring the freshness
CC of the fish product such as fresh, refrigerated, or frozen, fish food
CC product, and for screening compounds useful for treating stress in fish.
CC Nucleic acids encoding the antimicrobial peptides are useful for treating
CC stress in a fish. Sequences ABB07720-23 represent specific examples of
CC the antimicrobial peptides.
XX
SQ Sequence 22 AA;
XX
Query Match 100.0%; Score 121; DB 23; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.5e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FFHHIFRGIVHWGKTIHRLVTG 22
DB 1 FFHHIFRGIVHWGKTIHRLVTG 22
XX
RESULT 3
AAE20174
ID AAE20174 standard; peptide; 22 AA.
XX
AC AAE20174;
XX
DT 18-JUN-2002 (first entry)
XX
DE Fish mast cell antimicrobial peptide; piscidin 2 (P2).
XX
KM Antimicrobial peptide; endobiotic peptide; mast cell; piscidin; fish;
XX methicillin; vancomycin; streptogramin; microbial infection; stress;
XX bacterial disease; fish health; screening; gene therapy; antimicrobial;
XX tranquilizer; fungicide; protozoicide; antibacterial; antiparasitic.
XX
OS Morone chrysops.
XX
PN WO200214345-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US41696.
XX
PR 15-AUG-2000; 2000US-225354P.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Noga EJ, Silphaduang U;
XX
DR WPI; 2002-269176/31.
XX
PT Novel antimicrobial peptides (endobiotic peptides) isolated from mast
PT cells, specifically fish mast cells, useful for treating microbial
PT infections in humans or animals and for reducing antibiotic resistance
PT in bacteria -
XX
PS Claim 4; Page 18; 31pp; English.
XX
CC The invention relates to antimicrobial peptide (endobiotic peptide)
CC isolated from a mast cell. Such peptides obtained from fish mast cells
CC are referred as piscidins. The antimicrobial peptide is useful for
CC treating microbial infection in a subject, and for reducing antibiotic
CC (e.g., methicillin, vancomycin or streptogramin) resistance in bacteria
CC such as *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus aureus*,
CC *Klebsiella pneumoniae*, *Streptococcus faecalis* or *Shigella flexneri*. These
CC peptides are useful for treating microbial infections in human or animals
CC and for preparing a medicament for treating microbial infections. These
CC peptides are also useful for treating infections caused by microorganisms
CC other than bacteria, e.g., fungi, yeast, protozoa, parasite, etc. These
CC peptides are useful for treating stress or bacterial disease in fish and
CC monitoring fish health. The inverse relationship between endobiotic
CC levels and stress also allows assessment of freshness of fish food

CC product. Nucleic acid sequences encoding these peptides are useful for
CC diagnostic purposes, creating stress in a fish and for screening
CC additional endobiotic peptides. They are also useful in gene therapy.
CC The present sequence is piscidin 2 (P2) peptide isolated from the gills
CC of hybrid striped bass (Morone saxatilis x Morone chrysops).

XX Sequence 22 AA;

Query Match 97.5%; Score 118; DB 23; Length 22;
Best Local Similarity 95.5%; Pred. No. 1.9e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FFHHIFRGIVAGKTIHRLVTG 22
1 FFHHIFRGIVAGKTIHRLVTG 22
DB 1 FFHHIFRGIVAGKTIHRLVTG 22

RESULT 4

ABB07722
ID ABB07722 standard; peptide; 22 AA.
XX
AC ABB07722;
XX
DT 10-JUN-2002 (first entry)
XX
DE Antimicrobial peptide #3 from hybrid striped bass.
XX
KW Fish; hybrid striped bass; endobiotic; antibiotic; tranquilizer;
KM antimicrobial; antibacterial; fish food product.
XX
OS Morone saxatilis x Morone chrysops.
XX
PN WO200214346-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US41697.
XX
PR 15-AUG-2000; 2000US-225354P.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Noga EJ, Silphaduang U;
XX
DR WPI; 2002-269177/31.
XX
PT Novel antimicrobial compound (endobiotic peptide isolated from fish).
PT useful for monitoring fish health and monitoring freshness of fish food
PT product and for screening compounds which treat stress in fish -
XX
PS Claim 1; Page 16; 25pp; English.
XX
CC The invention provides antimicrobial peptides (ABB07720-23) isolated from
CC hybrid striped bass. The peptides were purified from the endobiotic
CC family. The endogenous endobiotic peptides exhibiting antimicrobial
CC activity) are useful for monitoring fish health, monitoring the freshness
CC of the fish product such as fresh, refrigerated, or frozen, fish food
CC product, and for screening compounds useful for treating stress in fish.
CC Nucleic acids encoding the antimicrobial peptides are useful for treating
CC stress in a fish. Sequences ABB07720-23 represent specific examples of
CC the antimicrobial peptides.

XX Sequence 22 AA;

Query Match 97.5%; Score 118; DB 23; Length 22;
Best Local Similarity 95.5%; Pred. No. 1.9e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FFHHIFRGIVAGKTIHRLVTG 22
1 FFHHIFRGIVAGKTIHRLVTG 22
DB 1 FFHHIFRGIVAGKTIHRLVTG 22

RESULT 5
AAE20172
ID AAE20172 standard; peptide; 22 AA.
XX
AC AAE20172;
XX
DT 18-JUN-2002 (first entry)
XX
DE Fish mast cell antimicrobial peptide, piscidin 3 (P3).
XX
KW Antimicrobial peptide; endobiotic peptide; mast cell; piscidin; fish;
KM metchnikellin; vancomycin; streptogramin; microbial infection; stress;
KW bacterial disease; fish health; screening; gene therapy; antimicrobial;
KW tranquilizer; fungicide; protozoicide; antibacterial; antiparasitic.
XX
OS Morone chrysops.
XX
PN WO200214345-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US41696.
XX
PR 15-AUG-2000; 2000US-225354P.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Noga EJ, Silphaduang U;
XX
DR WPI; 2002-269176/31.
XX
PT Novel antimicrobial peptides (endobiotic peptides) isolated from mast
PT cells, specifically fish mast cells, useful for treating microbial
PT infections in humans or animals and for reducing antibiotic resistance
PT in bacteria -
XX
XX
PS Claim 4; Page 17; 31pp; English.

XX
CC The invention relates to antimicrobial peptide (endobiotic peptide)
CC isolated from a mast cell. Such peptides obtained from fish mast cells
CC are referred to as piscidins. The antimicrobial peptide is useful for
CC treating microbial infection in a subject, and for reducing antibiotic
CC (e.g., metchnikellin, vancomycin or streptogramin) resistance in bacteria
CC such as *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus aureus*,
CC *Klebsiella pneumoniae*, *Streptococcus faecalis* or *Shigella flexneri*. These
CC peptides are useful for treating microbial infections in human or animals
CC and for preparing a medicament for treating microbial infections. These
CC peptides are also useful for treating infections caused by microorganisms
CC other than bacteria, e.g., fungi, yeast, protozoa, parasite, etc. These
CC peptides are useful for treating stress or bacterial disease in fish and
CC monitoring fish health. The inverse relationship between endobiotic
CC levels and stress also allows assessment of freshness of fish food
CC product. Nucleic acid sequences encoding these peptides are useful for
CC diagnostic purposes, treating stress in a fish and for screening
CC additional endobiotic peptides. They are also useful in gene therapy.
CC The present sequence is piscidin 3 (P3) peptide isolated from the gills
CC of hybrid striped bass (Morone saxatilis x Morone chrysops).

XX Sequence 22 AA;

Query Match 71.9%; Score 87; DB 23; Length 22;
Best Local Similarity 68.2%; Pred. No. 1.2e-06;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 FFHHIFRGIVAGKTIHRLVTG 22
1 FFHHIFRGIVAGKTIHRLVTG 22
DB 1 FFHHIFRGIVAGKTIHRLVTG 22

RESULT 6

ABB07720
ID ABB07720 standard; peptide; 22 AA.

```

XX AC ABB07720;
XX DT 10-JUN-2002 (first entry)
XX DE Antimicrobial peptide #1 from hybrid striped bass.
XX KW Fish; hybrid striped bass; endobiotic; antibiotic; tranquilizer;
XX KW antimicrobial; antibacterial; fish food product.
XX OS Morone saxatilis x Morone chrysops.
XX PN WO200214346-A2.
XX PD 21-FEB-2002.
XX PF 13-AUG-2001; 2001WO-US41697.
XX PR 15-AUG-2000; 2000US-225354P.
XX PA (UNNC-) UNIV NORTH CAROLINA STATE.
XX PI Noga EJ, Silphaduang U;
XX DR WPI; 2002-269177/31.
XX PS Claim 1; Page 16; 25pp; English.
XX CC The invention provides antimicrobial peptides (ABB07720-23) isolated from
CC hybrid striped bass. The peptides were purified from the endobiotic
CC family. The endogenous endobiotic peptides exhibiting antimicrobial
CC activity) are useful for monitoring fish health, monitoring the freshness
CC of the fish product such as fresh, refrigerated, or frozen, fish food
CC product, and for screening compounds useful for treating stress in fish.
CC Nucleic acids encoding the antimicrobial peptides are useful for treating
CC stress in a fish. Sequences ABB07720-23 represent specific examples of
CC the antimicrobial peptides.
XX SQ Sequence 22 AA;
XX
XX Query Match 71.9%; Score 87; DB 23; Length 22;
XX Best Local Similarity 68.2%; Pred. No. 1.2e-06;
XX Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 FFHHIFRGIVHWKTIHRLVTG 22
XX | ||||| ||||| : : : ||
XX Db 1 FIHHIFRGIVHAGRSIGRFLTG 22
XX
XX RESULT 7
XX AAE20177
XX ID AAE20177 standard; peptide; 24 AA.
XX AC AAE20177;
XX DT 18-JUN-2002 (first entry)
XX DE Antimicrobial peptide analogue.
XX KW Antimicrobial peptide; endobiotic peptide; mast cell; piscidin;
XX KW methicillin; vancomycin; streptomycin; microbial infection; stress;
XX KW bacterial disease; fish health; screening; gene therapy; antimicrobial;
XX KW tranquilizer; fungicide; protozoicide; antibacterial; antiparasitic.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Disulfide-bond 12..13
XX

```

```

PN WO200214345-A2.
XX PD 21-FEB-2002.
XX PF 13-AUG-2001; 2001WO-US41696.
XX PR 15-AUG-2000; 2000US-225354P.
XX PA (UNNC-) UNIV NORTH CAROLINA STATE.
XX PI Noga EJ, Silphaduang U;
XX DR WPI; 2002-269176/31.
XX PS Novel antimicrobial peptides (endobiotic peptides) isolated from mast
XX PT cells, specifically fish mast cells, useful for treating microbial
XX PT infections in humans or animals and for reducing antibiotic resistance
XX PT in bacteria -
XX
XX PS Example 9; Page 24; 31pp; English.
XX CC The invention relates to antimicrobial peptide (endobiotic peptide)
XX CC isolated from a mast cell. Such peptides obtained from fish mast cells
XX CC are referred as piscidins. The antimicrobial peptide is useful for
XX CC treating microbial infection in a subject, and for reducing antibiotic
XX CC (e.g., methicillin, vancomycin or streptomycin) resistance in bacteria
XX CC such as Pseudomonas aeruginosa, Escherichia coli, Staphylococcus aureus,
XX CC Klebsiella pneumoniae, Streptococcus faecalis or Shigella flexneri. These
XX CC peptides are useful for treating microbial infections in human or animals
XX CC and for preparing a medicament for treating microbial infections. These
XX CC peptides are also useful for treating infections caused by microorganisms
XX CC other than bacteria, e.g., fungi, yeast, protozoa, parasite, etc. These
XX CC peptides are useful for treating stress or bacterial disease in fish and
XX CC monitoring fish health. The inverse relationship between endobiotic
XX CC levels and stress also allows assessment of freshness of fish food
XX CC product. Nucleic acid sequences encoding these peptides are useful for
XX CC diagnostic purposes, treating stress in a fish and for screening
XX CC additional endobiotic peptides. They are also useful in gene therapy.
XX
XX SQ Sequence 24 AA;
XX
XX Query Match 53.7%; Score 65; DB 23; Length 24;
XX Best Local Similarity 100.0%; Pred. No. 0.0032;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 FFHHIFRGIVH 11
XX | ||||| |||||
XX Db 14 FFHHIFRGIVH 24
XX
XX RESULT 8
XX AAB96197
XX ID AAB96197 standard; Protein; 907 AA.
XX AC AAB96197;
XX DT 29-OCT-2001 (first entry)
XX DE Putative P. abyssi DNA-dependent RNA polymerase A subunit #1.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX OS Pyrococcus abyssi.
XX PN FR2792651-A1.
XX PD 27-OCT-2000.
XX PF 21-APR-1999; 99FR-0005034.
XX PR 21-APR-1999; 99FR-0005034.
XX

```

(CNRS) CNRS CENT NAT RECH SCI.
 (IFRE-) IFREMER INST FR RECH EXPL MER.
 Porterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O,
 Querellou J, Weissenbach J, Saurin W, Heilig R,
 WPI: 2001-126236/14.
 New nucleotide sequences isolated from *Pyrococcus abyssi* encode
 proteins useful in industry -
 Claim 7: Pages 838-841; 1657p; French.
 The present invention relates to the genomic sequence of *Pyrococcus*
abyssi (see AAF64431 and AAH41223-7) and *P. abyssi* is
 a hyperthermophilic archaeon, which is isolated from deep-sea
 hydrothermal vents. The present sequence is one such *P. abyssi* protein.
 The proteins of the present invention have various potential industrial
 uses, since the proteins are stable at very high temperatures, some up to
 110 degrees centigrade.
 Note: This patent is in the same patent family as WO2000065062, which
 contains additional sequences as shown in AAB9132-AA99143.
 AAH75903-AAH75920 and AAG6436.
 Sequence 907 AA;
 Query Match 40.5%; Score 49; DB 22; Length 907;
 Best Local Similarity 52.2%; Pred. No. 46;
 Matches 12; Conservative 4; Mismatches 3; Indels 4; Gaps 2;
 2 FHHI--FRGIVHG--KTIHRLV 20
 74 FGHILARVHVHGFATKTHRL 96
 RESULT 9
 AAB40499
 ID AAB40499 standard; Protein; 290 AA.
 AC AAB40499;
 DT 08-FEB-2001 (first entry)
 XX Human ORFX ORF263 polypeptide sequence SEQ ID NO:526.
 DE Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnery; antipapillary; antipapilloma; neoplastic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antipneumatic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive.
 XX Homo sapiens.
 OS
 XX
 XX MO200058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX

(CURA-) CURAGEN CORP.
 Shinkets RA, Leach M;
 WPI: 2000-602362/57.
 DR N-PSDB; AAC74708.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 Claim 11; Page 689; 5507p; English.
 AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
 CC antipapillary; antipapilloma; neoplastic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antipneumatic; antiviral; antifungal; antineumatic;
 CC anti-inflammatory; antibacterial; antipneumatic; antineumatic;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 290 AA;
 Query Match 40.1%; Score 48.5; DB 21; Length 290;
 Best Local Similarity 28.6%; Pred. No. 16;
 Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;
 1 FHHIIFR-----GIYHNGKTIHRLVT 21
 Db 167 FYGHFLREITVANGGFLTYGSEVHRMLT 194
 RESULT 10
 AAB85392
 ID AAB85392 standard; Protein; 392 AA.
 AC AAB85392;
 DT 17-SEP-2001 (first entry)
 XX Stem cell growth factor-like polypeptide.
 DE Stem cell growth factor-like polypeptide.
 KW Stem cell growth factor-like polypeptide; leukemia; hemophilia; human;
 KW degenerative disease; Alzheimer's disease; nutritional supplement;
 KW cytosolic; neuroprotective; hemostatic; antineoplastic; antineoplastic;
 KW gene-therapy; cell proliferation; stem cell growth factor.
 XX Homo sapiens.
 OS
 XX
 XX MO200153500-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX 23-DEC-2000; 2000WO-US35260.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 07-APR-2000; 2000US-0545714.
 PR 11-APR-2000; 2000US-0547358.
 XX

```
PA      (HYSE-) HYSEQ INC.  
XX  
XX Labat I, Tang YT, Drmanac RT, Liu C, Lee J, Mike NK, Childs J,  
PI Chao C;  
XI  
XN WPI; 2001-451909/48.  
DR N-PDSB; AAH23063.  
XX  
PS Example 2; Page 127-128; 154pp; English.

The invention provides novel human stem cell growth factor-like polypeptides and polynucleotides encoding them. The polypeptides having stem cell growth factor-like activity, can be expressed by standard recombinant methodology. The polynucleotides and polypeptides can be used to induce differentiation of embryonic and adult stem cells to give rise to different cell types. They may also be used in the treatment of leukemia, hemophilia, and degenerative diseases like Alzheimer's disease. They may also be utilized to generate new tissues and organs that may aid patients in need of transplants. They can also be used as nutritional supplements. The present sequence represents a stem cell growth factor-like polypeptide.



SQ Sequence 392 AA;



Query Match 40.1%; Score 48.5; DB 22; Length 392;  
Best Local Similarity 28.6%; Pred. No. 23;  
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1



OY 1 FFHHIFR-----GIWVGSKTIHLRVLT 21  
| : | | : | : | : | :  
DB 30 FYGHFLREITVATGCFIYTGEVVHRMLT 57



RESULT 11  
AAB85393  
ID AAB85393 standard; Protein; 392 AA.  
XX  
XX AC AAB85393;  
XX  
XX DT 17-SEP-2001 (first entry)  
XX  
XX DE Stem cell growth factor-like polypeptide.  
XX  
XX KM Degenerative factor-like polypeptide; leukemia; hemophilia; human;  
KM degenerative disease; Alzheimer's disease; nutritional supplement;  
KW cystostatic; nootropic; neuroprotective; hematostic; antiense-therapy;  
gene-therapy; cell proliferation; stem cell growth factor.  
XX  
OS Homo sapiens.  
XX  
XX PN WO200153500-A1.  
XX  
XX PD 26-JUL-2001.  
XX  
XX PE 23-DEC-2000; 2000WO-US35260.  
XX  
XX PR 21-JAN-2000; 2000US-0488725.  
PR 07-APR-2000; 2000US-0545714.  
PR 11-APR-2000; 2000US-0547358.  
XX  
XX PA (HYSE-) HYSEQ INC.



Labat I, Tang YT, Drmanac RT, Liu C, Lee J, Mike NK, Childs J,  
PI Chao C;  
XI  
XN WPI; 2001-451909/48.  
DR N-PDSB; AAH23064, AAH23065.



Isolated polypeptide with stem cell growth factor-like activity for


```

```
PT treatment of leukemia, hemophilia, and degenerative diseases like
PT Alzheimer's disease and to generate new tissues and organs -
PS Claim 10; Page 129-133; 154pp; English.
XX
CC The invention provides novel human stem cell growth factor-like
CC polypeptides and polynucleotides encoding them. The polypeptides having
CC stem cell growth factor-like activity, can be expressed by standard
CC recombinant methodology. The polynucleotides and polypeptides can be
CC used to induce differentiation of embryonic and adult stem cells to give
CC rise to different cell types. They may also be used in the treatment of
CC leukemia, hemophilia, and degenerative diseases like Alzheimer's disease.
CC They may also be utilized to generate new tissues and organs that aid
CC patients in need of transplants. They can also be used as nutritional
CC supplements. The present sequence represents a stem cell growth factor-
CC like polypeptide.
SQ Sequence 392 AA;

Query Match 40.1%; Score 48.5; DB 22; Length 392;
Best Local Similarity 28.6%; Pred. No. 23;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1.

Qy 1 FFHAIER-----GIHWGKTIHLRVT 21
| : | | | : | : | : | : | : |
Db 30 FYGHFLREITVATGGRITYTGEVHHMLT 57

RESULT 12
AAG64527
ID AAG64527 standard; Protein; 425 AA.
XX
AC AAG64527;
XX
DI 04-OCT-2001 (first entry)
XX
DE Human polypeptide 784_3137 contig SEQ ID NO 2.
XX
KW Human; antisense-therapy; gene-therapy; diagnostic; forensic;
gene mapping.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 287 /note= "Encoded by TCTT"
FT Misc-difference 293 /note= "Encoded by CCATAG"
FT Misc-difference 302 /note= "Encoded by CCTATA"
FT Misc-difference 307 /note= "Encoded by TCCCTAA"
FT FT /note= "Encoded by GA"
XX
WO200152616-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000MO-US$3190.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 23-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
DR WPI; 2001-451890/48.
N-PSDB; AAAH5398.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
antibodies and research use -
```

PS Claim 1; Page 15; 135pp; English.

XX The invention relates to an isolated human polynucleotide (AAH75398)

CC encoding a novel polypeptide (AAG64527) useful in antisense-therapy and

CC gene-therapy, in diagnostics, forensics, gene mapping and identification

CC of mutations responsible for genetic disorders and other traits.

CC Polynucleotide sequences with potential homology were also identified

CC (AAH93283-AAH93356).

XX

SQ Sequence 425 AA;

Query Match 40.1%; Score 48.5; DB 22; Length 425;

Best Local Similarity 28.6%; Pred. No. 25;

Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

1 FFHHIFR-----GIVHGKTIHRLVT 21

65 FYGHFLREITVATGSPITGSEVVRHMLT 92

Db

RESULT 13

AAH85399

ID AAB85399 standard; Protein: 425 AA.

XX

AC AAB85399;

XX

DT 17-SEP-2001 (first entry)

XX

DE Stem cell growth factor-like polypeptide fragment.

XX

KW Stem cell growth factor-like polypeptide; leukemia; hemophilia; human;

KW degenerative disease; Alzheimer's disease; nutritional supplement;

KW cytoskeletal; neurotropic; neuroprotective; hemostatic; antisense-therapy;

KW gene-therapy; cell proliferation; stem cell growth factor.

XX

OS Homo sapiens.

XX

PN WO200153500-A1.

XX

PD 26-JUL-2001.

XX

PF 23-DEC-2000; 2000WO-US35260.

XX

PR 21-JAN-2000; 2000US-0488725.

PR 07-APR-2000; 2000US-0545714.

PR 11-APR-2000; 2000US-0547358.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Labat I, Tang YT, Drmanac RT, Liu C, Lee J, Mize NK, Childs J;

PI Chao C;

XX

DR WPI; 2001-451908/48.

XX

PT isolated polypeptide with stem cell growth factor-like activity for

PT treatment of leukemia, hemophilia, and degenerative diseases like

PT Alzheimer's disease and to generate new tissues and organs -

XX

PS Claim 10; Page 147-149; 154pp; English.

XX

CC The invention provides novel human stem cell growth factor-like

CC polypeptides and polynucleotides encoding them. The polypeptides having

CC stem cell growth factor-like activity, can be expressed by standard

CC recombinant methodology. The polynucleotides and polypeptides can be

CC used to induce differentiation of embryonic and adult stem cells to give

CC rise to different cell types. They may also be used in the treatment of

CC leukemia, hemophilia, and degenerative diseases like Alzheimer's disease.

CC They may also be utilized to generate new tissues and organs that may aid

CC patients in need of transplants. They can also be used as nutritional

CC supplements. The present sequence represents a stem cell growth factor-

CC like polypeptide fragment.

XX

SQ Sequence 425 AA;

Query Match 40.1%; Score 48.5; DB 22; Length 425;

Best Local Similarity 28.6%; Pred. No. 25;

Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

1 FFHHIFR-----GIVHGKTIHRLVT 21

65 FYGHFLREITVATGSPITGSEVVRHMLT 92

Db

RESULT 14

AAH40853

ID AAM40853 standard; Protein: 427 AA.

XX

AC AAM40853;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 5784.

XX

KW Human; neurotropic; immunosuppressant; cytoskeletal; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US34263.

XX

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Aarndt V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Weinman T, Xu C, Xue AD, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX

DR WPI; 2001-442253/47.

DR N-PSDB; AAL60009.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX

PS Example 2; SEQ ID NO 5784; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAM38642-AA44213) with neurotropic,

CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.
XX
SQ Sequence 427 AA;
Query Match 40.1%; Score 48.5; DB 22; Length 427;
Best Local Similarity 28.6%; Pred. No. 25;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;
QY 1 FFHHIFR-----GIVHVKTIHRLVT 21
DB 65 FYGHFLREITVATGGRFYTGVEVHRLMT 92
RESULT 15
AAM40854
ID AAM40854 standard; Procein; 427 AA.
XX
AC AAM40854;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5785.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia.
XX
OS Homo sapiens.
XX
PN MO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AAI60010.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5785; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 427 AA;
Query Match 40.1%; Score 48.5; DB 22; Length 427;
Best Local Similarity 28.6%; Pred. No. 25;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;
QY 1 FFHHIFR-----GIVHVKTIHRLVT 21
DB 65 FYGHFLREITVATGGRFYTGVEVHRLMT 92
Search completed: May 23, 2003, 08:55:23
Job time : 32 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:54:25 ; Search time 11.5 Seconds
(without alignments)
56.287 Million cell updates/sec

Title: US-09-929-788-2
Perfect score: 121
Sequence: 1 FFHHFRGIVHGKTHRLVVG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCtus.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	39.7	436	1 US-08-674-168-6	Sequence 6, Appl
2	44	36.4	436	1 US-08-146-010A-2	Sequence 2, Appl
3	44	36.4	436	1 US-08-674-168-2	Sequence 2, Appl
4	43	35.5	399	2 US-08-834-655-4	Sequence 4, Appl
5	43	35.5	399	3 US-08-834-033A-4	Sequence 4, Appl
6	43	35.5	399	4 US-09-363-574-4	Sequence 4, Appl
7	43	35.5	399	4 US-09-363-574-4	Sequence 4, Appl
8	43	35.5	399	4 US-09-330-235-20	Sequence 20, Appl
9	41	33.9	466	4 US-09-134-001C-3526	Sequence 3526, Ap
10	41	33.9	471	4 US-09-134-001C-4904	Sequence 4904, Ap
11	41	33.9	506	4 US-09-370-838-34	Sequence 34, Appl
12	41	33.9	764	4 US-09-641-741-31	Sequence 31, Appl
13	40	33.1	23	2 US-08-808-277A-13	Sequence 13, Appl
14	40	33.1	345	2 US-08-758-621-14	Sequence 14, Appl
15	40	33.1	345	2 US-09-107-858-14	Sequence 14, Appl
16	40	33.1	810	2 US-08-820-170A-34	Sequence 34, Appl
17	40	33.1	810	3 US-09-055-699-34	Sequence 34, Appl
18	40	33.1	810	4 US-09-273-565-34	Sequence 34, Appl
19	40	33.1	810	4 US-09-565-538-34	Sequence 34, Appl
20	40	33.1	810	4 US-09-661-468-34	Sequence 34, Appl
21	39	32.2	151	2 US-08-858-767-30	Sequence 30, Appl
22	39	32.2	151	2 US-08-863-028-30	Sequence 30, Appl
23	39	32.2	223	4 US-08-961-083-162	Sequence 162, App
24	39	32.2	223	4 US-08-147-772-2	Sequence 2, Appl
25	39	32.2	288	2 US-08-456-104-6	Sequence 6, Appl
26	39	32.2	288	2 US-08-101-624-23	Sequence 2, Appl
27	39	32.2	288	2 US-08-751-767A-6	Sequence 6, Appl

28	39	32.2	288	3 US-08-153-262-2	Sequence 2, Appl
29	39	32.2	288	3 US-08-479-744A-29	Sequence 29, Appl
30	39	32.2	288	4 US-08-280-757B-29	Sequence 2, Appl
31	39	32.2	288	4 US-09-159-135-2	Sequence 2, Appl
32	39	32.2	288	4 US-08-205-697A-19	Sequence 19, Appl
33	39	32.2	288	4 US-08-702-525-19	Sequence 19, Appl
34	39	32.2	288	4 US-09-450-798-2	Sequence 2, Appl
35	39	32.2	288	4 US-08-403-253A-2	Sequence 2, Appl
36	39	32.2	288	4 US-09-651-200-13	Sequence 13, Appl
37	39	32.2	288	4 US-09-651-200-14	Sequence 14, Appl
38	39	32.2	288	5 PCT-US95-02576-19	Sequence 19, Appl
39	39	32.2	371	2 US-08-837-593-6	Sequence 6, Appl
40	39	32.2	416	3 US-08-554-385-21	Sequence 21, Appl
41	39	32.2	473	4 US-09-171-945-131	Sequence 131, App
42	39	32.2	478	4 US-09-004-838-51	Sequence 51, Appl
43	39	32.2	1074	4 US-09-004-838-111	Sequence 111, App
44	38.5	31.8	23	2 US-08-808-277A-18	Sequence 18, Appl
45	38.5	31.8	170	3 US-09-130-663-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-674-168-6
Sequence 6, Application US/08674168
Patent No. 5804414
GENERAL INFORMATION:
APPLICANT: MORIYA, Mika
APPLICANT: MATSUI, Hiroshi
APPLICANT: YOKOZAKI, Kenzo
APPLICANT: HIRANO, Seiko
APPLICANT: HAYAKAWA, Atsushi
APPLICANT: IZUI, Masako
APPLICANT: SUGIMOTO, Masakazu
TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
TITLE OF INVENTION: ARTIFICIAL TRANSPOSON
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSER: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSER: P. C.
STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/674,168
FILING DATE: 01-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-166541
FILING DATE: 30-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-810-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-674-168-6

Query Match 39.7% Score 48; DB 1; Length 436;
Best Local Similarity 35.3% Pred. No. 6.3;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HHIFRGIVHVGKTIHRL 19
Db 417 HYLRLCLHSGQLTHKI 433

RESULT 2
US-08-146-010A-2
Sequence 2, Application US/08146010A
Patent No. 5591577

GENERAL INFORMATION:

APPLICANT: TSUCHIYA, MAKOTO

APPLICANT: MORIYA, MIKO

APPLICANT: MIWA, KIYOSHI

TITLE OF INVENTION: MOBILE GENETIC ELEMENT ORIGINATED FROM

TITLE OF INVENTION: BREVIBACTERIUM STRAIN

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/146,010A

FILING DATE: 12-NOV-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 52694/92

FILING DATE: 11-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-649-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 436 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-146-010A-2

Query Match 36.4% Score 44; DB 1; Length 436;
Best Local Similarity 35.3% Pred. No. 27;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HHIFRGIVHVGKTIHRL 19
Db 417 HYLRLCLHSGQLTHKI 433

RESULT 3
US-08-674-168-2

Sequence 2, Application US/08674168

Patent No. 5804414

GENERAL INFORMATION:

APPLICANT: MORIYA, MIKA

APPLICANT: MATSUI, HITOSHI

APPLICANT: YOKOZEKI, KENZO

APPLICANT: HIRANO, Seiko
APPLICANT: HAYAKAWA, Atsushi
APPLICANT: IZUI, Masako
APPLICANT: SUGIMOTO, Masakazu
TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
TITLE OF INVENTION: ARTIFICIAL TRANSPOSON
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/674,168

FILING DATE: 01-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-166541

FILING DATE: 30-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-810-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 436 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-674-168-2

Query Match 36.4% Score 44; DB 1; Length 436;
Best Local Similarity 35.3% Pred. No. 27;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HHIFRGIVHVGKTIHRL 19
Db 417 HYLRLCLHSGQLTHKI 433

RESULT 4
US-08-834-655-4

Sequence 4, Application US/08834655

Patent No. 5968809

GENERAL INFORMATION:

APPLICANT: KUTZON, DEBORAH

APPLICANT: MURKERT, PRADIP

APPLICANT: HUNG, YUNG-SHENG

APPLICANT: THURMOND, JENNIFER

APPLICANT: CHAUDHARY, SUNITA

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS

TITLE OF INVENTION: OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS IN PLANTS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: RAE-VENTER LAW GROUP, P.C.

STREET: 260 SHERIDAN AVENUE, P.O. BOX 60039

CITY: PALO ALTO

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

RESULT 7
US-09-363-526-4

```
; Sequence 4, Application US/09363526
; Patent No. 6410288
; GENERAL INFORMATION:
; APPLICANT: KUTZON, DEBORAH
; APPLICANT: MOKERJI, PRADIP
; APPLICANT: HUANG, YUNG-SHENG
; APPLICANT: THURMOND, JENNIFER
; APPLICANT: CHAUDHARY, SUNITA
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS
; TITLE OF INVENTION: OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS IN PLANTS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESS: LIMBACH AND LIMBACH L.L.P.
; STREET: 2001 FERRY BUILDING
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/363,526
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MARD, MICHAEL R.
; REGISTRATION NUMBER: 38,651
; REFERENCE/DOCKET NUMBER: CGAB-201 USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 433-4150
; TELEFAX: (415) 433-8716
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-363-526-4

Query Match          35.5%; Score 43; DB 4; Length 399;
Best Local Similarity 47.1%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVKTIH 17
DB 326 FLDHMFHGIHVHVAHH 342

RESULT 8
US-09-330-235-20
; Sequence 20, Application US/09330235
; Patent No. 6459018
; GENERAL INFORMATION:
; APPLICANT: Kutzon, Debbie
; TITLE OF INVENTION: POLYUNSATURATED FATTY ACIDS IN PLANTS
; FILE REFERENCE: MOCO.156.00US
; CURRENT APPLICATION NUMBER: US/09/330,235
; CURRENT FILING DATE: 1999-06-10
; PRIOR APPLICATION NUMBER: 60/089,043
; PRIOR FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent version 3.0
; SEQ ID NO 20
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Mortierella alpina
US-09-330-235-20
```

```
Query Match          35.5%; Score 43; DB 4; Length 399;
Best Local Similarity 47.1%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVKTIH 17
DB 326 FLDHMFHGIHVHVAHH 342
```

```
RESULT 9
US-09-134-001C-3526
; Sequence 3526, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3526
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3526
```

```
Query Match          33.9%; Score 41; DB 4; Length 466;
Best Local Similarity 33.3%; Pred. No. 85;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 2 FFHHIFRGIVHVKTIHRLVGTG 22
DB 47 FYHVDMSIVNIATVSLTLTG 67
```

```
RESULT 10
US-09-134-001C-4904
; Sequence 4904, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4904
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4904
```

```
Query Match          33.9%; Score 41; DB 4; Length 471;
Best Local Similarity 70.0%; Pred. No. 86;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 IFRGIVHVK 14
DB 361 IYRGIVHVKGR 370
```

```
RESULT 11
US-09-370-838-34
```


APPLICATION NUMBER: US/08/758,621
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/018,578
FILING DATE: 29-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silveri, Jean M.
REGISTRATION NUMBER: 39,030
REFERENCE/DOCKET NUMBER: DCI-099CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-758-621-14

Query Match 33.1%; Score 40; DB 2; Length 345;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FHHIRGI 9
|||: ||
Db 229 FHHIRGI 236

RESULT 15
US-09-107-858-14
Sequence 14, Application US/09107858
Patent No. 6162900
GENERAL INFORMATION:
APPLICANT: Guerinet, Mary Lou et al.
TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR
FILE REFERENCE: DCI-099CPDV
CURRENT APPLICATION NUMBER: US/09/107,858
CURRENT FILING DATE: 1998-06-30
EARLIER APPLICATION NUMBER: 08/758,621
EARLIER FILING DATE: 1996-11-27
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 345
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-107-858-14

Query Match 33.1%; Score 40; DB 4; Length 345;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FHHIRGI 9
|||: ||
Db 229 FHHIRGI 236

Search completed: May 23, 2003, 08:57:46
Job time : 13.5 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:56:50 ; Search time 16 Seconds
(without alignments)

136.362 Million cell updates/sec

Title: US-09-929-788-2

Perfect score: 121

Sequence: 1 FFHHIFRGIVHGKTIHRLVTG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121	100.0	22	US-09-929-788-2	Sequence 2, Appli
2	118	97.5	22	US-09-929-788-3	Sequence 3, Appli
3	114	94.2	23	US-10-076-816-42	Sequence 42, Appli
4	87	71.9	22	US-09-929-788-1	Sequence 1, Appli
5	48.5	40.1	118	US-09-867-550-198	Sequence 198, App
6	48.5	40.1	332	US-09-912-935-23	Sequence 23, Appli
7	48.5	40.1	332	US-09-912-935-25	Sequence 25, Appli
8	48.5	40.1	425	US-09-912-935-35	Sequence 35, Appli
9	48.5	40.1	449	US-09-912-935-34	Sequence 34, Appli
10	48.5	40.1	529	US-09-912-935-31	Sequence 31, Appli
11	48.5	40.1	529	US-10-066-500-128	Sequence 128, App
12	48.5	40.1	529	US-10-174-590-472	Sequence 472, App
13	48.5	40.1	529	US-10-176-758-472	Sequence 472, App
14	48.5	40.1	529	US-10-175-737-472	Sequence 472, App
15	48.5	40.1	529	US-09-912-935-28	Sequence 28, Appli
16	48.5	40.1	529	US-09-912-935-40	Sequence 40, Appli
17	48.5	40.1	529	US-10-173-706-472	Sequence 472, App
18	48.5	40.1	529	US-10-175-738-472	Sequence 472, App
19	48.5	40.1	529	US-10-175-752-472	Sequence 472, App

20	48.5	40.1	529	US-10-176-482-472	Sequence 472, App
21	48.5	40.1	529	US-10-176-757-472	Sequence 472, App
22	48.5	40.1	529	US-10-176-913-472	Sequence 472, App
23	48.5	40.1	529	US-10-180-552-472	Sequence 472, App
24	48.5	40.1	529	US-10-180-557-472	Sequence 472, App
25	48.5	40.1	529	US-10-173-700-472	Sequence 472, App
26	48.5	40.1	529	US-10-174-572-472	Sequence 472, App
27	48.5	40.1	529	US-10-174-579-472	Sequence 472, App
28	48.5	40.1	529	US-10-174-582-472	Sequence 472, App
29	48.5	40.1	529	US-10-174-588-472	Sequence 472, App
30	48.5	40.1	529	US-10-175-739-472	Sequence 472, App
31	48.5	40.1	529	US-10-175-740-472	Sequence 472, App
32	48.5	40.1	529	US-10-175-743-472	Sequence 472, App
33	48.5	40.1	529	US-10-176-488-472	Sequence 472, App
34	48.5	40.1	529	US-10-176-492-472	Sequence 472, App
35	48.5	40.1	529	US-10-176-747-472	Sequence 472, App
36	48.5	40.1	529	US-10-176-750-472	Sequence 472, App
37	48.5	40.1	529	US-10-176-985-472	Sequence 472, App
38	48.5	40.1	529	US-10-176-987-472	Sequence 472, App
39	48.5	40.1	529	US-10-176-991-472	Sequence 472, App
40	48.5	40.1	529	US-10-176-992-472	Sequence 472, App
41	48.5	40.1	529	US-10-176-993-472	Sequence 472, App
42	48.5	40.1	529	US-10-184-658-472	Sequence 472, App
43	48.5	40.1	529	US-10-062-796-128	Sequence 128, App
44	48.5	40.1	529	US-10-066-273-128	Sequence 128, App
45	48.5	40.1	529	US-10-066-494-128	Sequence 128, App

ALIGNMENTS

RESULT 1
US-09-929-788-2
Sequence 2, Application US/09929788
Publication No. US20030083247A1
GENERAL INFORMATION:
APPLICANT: No. US20030083247A1a, Edward J.
APPLICANT: Sijphaduang, Umaphorn
TITLE OR INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
FILE REFERENCE: 5051.519
CURRENT APPLICATION NUMBER: US/09/929, 788
CURRENT FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: US 60/225,354
PRIOR FILING DATE: 2000-08-15
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 22
TYPE: PRT
ORGANISM: Morone saxatilis x Morone chrysops
US-09-929-788-2

Query Match 100.0%; Score 121; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 7, 1e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHGKTIHRLVTG 22
DB 1 FFHHIFRGIVHGKTIHRLVTG 22

RESULT 2
US-09-929-788-3
Sequence 3, Application US/09929788
Publication No. US20030083247A1
GENERAL INFORMATION:
APPLICANT: No. US20030083247A1a, Edward J.
APPLICANT: Sijphaduang, Umaphorn
TITLE OR INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
FILE REFERENCE: 5051.519
CURRENT APPLICATION NUMBER: US/09/929, 788
CURRENT FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: US 60/225,354

;; PRIOR FILING DATE: 2000-08-15
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 3
;; LENGTH: 22
;; TYPE: PRT
;; ORGANISM: Morone saxatilis x Morone chrysops
US-09-929-788-3

Query Match 97.5%; Score 118; DB 9; Length 22;
Best Local Similarity 95.5%; Pred. No. 2e-11;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVGKTIHRLVTG 22
DB 1 FFHHIFRGIVHVGKTIHRLVTG 22

RESULT 3
US-10-076-816-42
; Sequence 42, Application US/10076816
; Publication No. US20030056244A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Ning
; APPLICANT: Rodriguez, Raymond
; APPLICANT: Hagie, Frank E.
; TITLE OF INVENTION: Feed Additive Compositions and Methods
; FILE REFERENCE: 50665-8021.US00
; CURRENT APPLICATION NUMBER: US/10/076,816
; PRIOR FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/269,188
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/847,232
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/266,929
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: US 60/201,182
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Morone sp.
US-10-076-816-42

Query Match 94.2%; Score 114; DB 9; Length 23;
Best Local Similarity 95.5%; Pred. No. 8.4e-11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVGKTIHRLVTG 22
DB 1 FFHHIFRGIVHVGKTIHRLVTG 22

RESULT 4
US-09-929-788-1
; Sequence 1, Application US/09929788
; Publication No. US20030083247A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030083247A1a, Edward J.
; APPLICANT: Silphaduang, Umaporn
; TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES ISOLATED FROM FISH
; FILE REFERENCE: 5051.519
; CURRENT APPLICATION NUMBER: US/09/929,788
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,354
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 22
; TYPE: PRT

;; ORGANISM: Morone saxatilis x Morone chrysops
US-09-929-788-1

Query Match 71.9%; Score 87; DB 9; Length 22;
Best Local Similarity 68.2%; Pred. No. 9.4e-07;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVGKTIHRLVTG 22
DB 1 FFHHIFRGIVHAGRSIGRFLTG 22

RESULT 5
US-09-867-550-198
; Sequence 198, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1e1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 198
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-198

Query Match 40.1%; Score 48.5; DB 10; Length 118;
Best Local Similarity 28.6%; Pred. No. 3.6;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

QY 1 FFHHIFR-----GIVHVGKTIHRLVT 21
DB 53 FYGHFLRRLTIVATGCFYTGEBVYHRLMT 80

RESULT 6
US-09-912-935-23
; Sequence 23, Application US/09912935
; Publication No. US2003002825A1
; GENERAL INFORMATION:
; APPLICANT: Nishikawa, Mitsuo et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
; FILE REFERENCE: 32066/37483
; CURRENT APPLICATION NUMBER: US/09/912,935
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: PCT/US00/35260
; PRIOR FILING DATE: 2000-12-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-912-935-23

Query Match 40.1%; Score 48.5; DB 9; Length 392;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

QY 1 FFHHIFR-----GIVHVGKTIHRLVT 21

Db 30 FYGHFLREITVATGFGFYTGGEVVRMLT 57

RESULT 7

US-09-912-935-25
; Sequence 25, Application US/09912935
; Publication No. US20030022825A1

GENERAL INFORMATION:

APPLICANT: Nishikawa, Mitsuo et al.

TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
POLYPEPTIDES AND POLYNUCLEOTIDES

FILE REFERENCE: 32066/37483

CURRENT APPLICATION NUMBER: US/09/912,935

CURRENT FILING DATE: 2001-07-24

PRIOR APPLICATION NUMBER: PCT/US00/35260

PRIOR FILING DATE: 2000-12-23

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn version 3.0

SEQ ID NO 25

LENGTH: 392

TYPE: PRT

ORGANISM: Homo sapiens

US-09-912-935-25

Query Match 40.1%; Score 48.5; DB 9; Length 392;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

QY 1 FFHHIFR-----GIVHVGKTIHRLVT 21

Db 30 FYGHFLREITVATGFGFYTGGEVVRMLT 57

RESULT 8

US-09-912-935-35

; Sequence 35, Application US/09912935

; Publication No. US20030022825A1

GENERAL INFORMATION:

APPLICANT: Nishikawa, Mitsuo et al.

TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
POLYPEPTIDES AND POLYNUCLEOTIDES

FILE REFERENCE: 32066/37483

CURRENT APPLICATION NUMBER: US/09/912,935

CURRENT FILING DATE: 2001-07-24

PRIOR APPLICATION NUMBER: PCT/US00/35260

PRIOR FILING DATE: 2000-12-23

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn version 3.0

SEQ ID NO 35

LENGTH: 425

TYPE: PRT

ORGANISM: Homo sapiens

US-09-912-935-35

Query Match 40.1%; Score 48.5; DB 9; Length 425;
Best Local Similarity 28.6%; Pred. No. 15;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

QY 1 FFHHIFR-----GIVHVGKTIHRLVT 21

Db 65 FYGHFLREITVATGFGFYTGGEVVRMLT 92

RESULT 9

US-09-912-935-34

; Sequence 34, Application US/09912935

; Publication No. US20030022825A1

GENERAL INFORMATION:

APPLICANT: Nishikawa, Mitsuo et al.

TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
POLYPEPTIDES AND POLYNUCLEOTIDES

FILE REFERENCE: 32066/37483

CURRENT APPLICATION NUMBER: US/09/912,935

; CURRENT FILING DATE: 2001-07-24

; PRIOR APPLICATION NUMBER: PCT/US00/35260

; PRIOR FILING DATE: 2000-12-23

; NUMBER OF SEQ ID NOS: 53

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 34

; LENGTH: 449

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-912-935-34

Query Match 40.1%; Score 48.5; DB 9; Length 449;
Best Local Similarity 28.6%; Pred. No. 15;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

QY 1 FFHHIFR-----GIVHVGKTIHRLVT 21

Db 167 FYGHFLREITVATGFGFYTGGEVVRMLT 194

RESULT 10

US-09-912-935-31

; Sequence 31, Application US/09912935

; Publication No. US20030022825A1

GENERAL INFORMATION:

APPLICANT: Nishikawa, Mitsuo et al.

TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO STEM CELL GROWTH FACTOR-LIKE
POLYPEPTIDES AND POLYNUCLEOTIDES

FILE REFERENCE: 32066/37483

CURRENT APPLICATION NUMBER: US/09/912,935

CURRENT FILING DATE: 2001-07-24

PRIOR APPLICATION NUMBER: PCT/US00/35260

PRIOR FILING DATE: 2000-12-23

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn version 3.0

SEQ ID NO 31

LENGTH: 499

TYPE: PRT

ORGANISM: Homo sapiens

US-09-912-935-31

Query Match 40.1%; Score 48.5; DB 9; Length 499;
Best Local Similarity 28.6%; Pred. No. 17;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

QY 1 FFHHIFR-----GIVHVGKTIHRLVT 21

Db 137 FYGHFLREITVATGFGFYTGGEVVRMLT 164

RESULT 11

US-10-066-500-128

; Sequence 128, Application US/10066500

; Patent No. US20020177165A1

GENERAL INFORMATION:

APPLICANT: Avi J. Ashkenazi

APPLICANT: Kevin P. Baker

APPLICANT: David A. Bockstein

APPLICANT: Luc Desnoyers

APPLICANT: Dan L. Baton

APPLICANT: Napoleone Ferrara

APPLICANT: Sherman Fong

APPLICANT: Wei-Qiang Gao

APPLICANT: Hanspeter Geiber

APPLICANT: Mary E. Gerltisen

APPLICANT: Audrey Goddard

APPLICANT: Paul J. Godowski

APPLICANT: Austin L. Gurney

APPLICANT: Ivar J. Kjaevlin

APPLICANT: Jennie P. Mather

APPLICANT: Mary A. Napier

APPLICANT: James Pan

APPLICANT: Nicholas F. Paoni

APPLICANT: Margaret Ann Roy
APPLICANT: Timothy A. Stewart
APPLICANT: Daniel Tumas
APPLICANT: Colin K. Watanabe
APPLICANT: P. Mickey Williams
APPLICANT: William I. Wood
APPLICANT: Zemin Zang
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P130R1C7
CURRENT FILING DATE: 2002-02-01
PRIOR APPLICATION NUMBER: 10/002,796
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/06364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066840
PRIOR FILING DATE: 1997-11-25
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/095998
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/097000
PRIOR FILING DATE: 1998-08-18
PRIOR APPLICATION NUMBER: 60/099601
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100858
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106032
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/109304
PRIOR FILING DATE: 1998-11-20
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/139695
PRIOR FILING DATE: 1999-06-15
PRIOR APPLICATION NUMBER: 60/145070
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698

PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/149396
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 08/918874
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 08/933821
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 08/960507
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 09/114844
PRIOR FILING DATE: 1998-07-14
PRIOR APPLICATION NUMBER: 09/136801
PRIOR FILING DATE: 1998-08-19
PRIOR APPLICATION NUMBER: 09/136804
PRIOR FILING DATE: 1998-08-19
PRIOR APPLICATION NUMBER: 09/136828
PRIOR FILING DATE: 1998-08-19
PRIOR APPLICATION NUMBER: 09/158342
PRIOR FILING DATE: 1998-09-21
PRIOR APPLICATION NUMBER: 09/180997
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 09/202088
PRIOR FILING DATE: 1998-12-08
PRIOR APPLICATION NUMBER: 09/25311
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: 09/254460
PRIOR FILING DATE: 1999-03-09
PRIOR APPLICATION NUMBER: 09/254465
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: 09/284663
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 09/332928
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/332929
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/333075
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/333077
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380139
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/403296
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 09/423844
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 09/522342
PRIOR FILING DATE: 2000-03-09
PRIOR APPLICATION NUMBER: 09/548815
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 09/664610
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: 09/665350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: 09/709238
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 09/767609
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 09/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 09/808689
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 09/866028
PRIOR FILING DATE: 2001-05-25

```

; PRIOR APPLICATION NUMBER: 09/870574
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 09/872035
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 09/886342
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: PCT/US98/14552
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: PCT/US98/18824
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US98/19093
; PRIOR FILING DATE: 1998-09-14
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: PCT/US98/19437
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: PCT/US98/24855
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: 1998-12-01
; PRIOR APPLICATION NUMBER: PCT/US98/25190
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: PCT/US99/05028
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547

Query Match          40.1%; Score 48.5; DB 9; Length 529;
Best Local Similarity 28.6%; Pred. No. 18;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

```

```

QY 1 FFHHIFR-----GIVHVKTIHRLVT 21
Db 167 FYGHFLREITVATGFIYTGVEVHRMLT 194

```

```

RESULT 12
US-10-174-590-472
; Sequence 472, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; PRIOR APPLICATION: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 472
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-472

```

Query Match 40.1%; Score 48.5; DB 9; Length 529;

```

Best Local Similarity 28.6%; Pred. No. 18;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

```

```

QY 1 FFHHIFR-----GIVHVKTIHRLVT 21
Db 167 FYGHFLREITVATGFIYTGVEVHRMLT 194

```

```

RESULT 13
US-10-176-758-472
; Sequence 472, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; PRIOR APPLICATION: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 472
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-472

```

```

Query Match          40.1%; Score 48.5; DB 9; Length 529;
Best Local Similarity 28.6%; Pred. No. 18;
Matches 8; Conservative 7; Mismatches 6; Indels 7; Gaps 1;

```

```

QY 1 FFHHIFR-----GIVHVKTIHRLVT 21
Db 167 FYGHFLREITVATGFIYTGVEVHRMLT 194

```

```

RESULT 14
US-10-175-737-472
; Sequence 472, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; PRIOR APPLICATION: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 472
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo Sapien

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:53:20 ; Search time 14.Seconds
(without alignments)
151.068 Million cell updates/sec

Title: US-09-929-788-2

Sequence: 1 FFHHIFRGIVHVGKTIHRLVYG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	40.5	907	2 H71031	probable DNA-directed
2	49	40.5	907	2 B75182	DNA-directed RNA p
3	48	39.7	351	2 AD0273	probable integral
4	48	39.7	404	2 S64944	hypothetical prote
5	48	39.7	436	2 S43613	transposase (inser
6	47	38.8	476	2 G84634	probable prollycar
7	47	38.8	612	2 AC1208	GTP-binding elonga
8	47	38.8	612	2 ARI564	GTP-binding elonga
9	47	38.8	905	1 S25564	DNA-directed RNA p
10	47	38.8	952	2 C82309	probable insulinas
11	46	38.0	173	2 B69354	ferritin homolog -
12	46	38.0	372	2 T52621	mitogen-activated
13	45	37.2	201	2 T36329	probable GTP cyclo
14	45	36.4	201	2 JC4742	transposase - Cory
15	44	36.4	513	2 C75553	hypothetical prote
16	44	36.4	917	1 VGBEBH	glycoprotein B pre
17	44	36.4	2550	2 VGB345	vesicular transpor
18	43	35.5	192	2 A42116	repressor for cyto
19	43	35.5	299	2 T47003	hypothetical prote
20	43	35.5	326	2 AF0238	probable acetyltra
21	43	35.5	326	2 B40141	mitochondrial solu
22	43	35.5	330	2 S26596	Graves disease mlt
23	43	35.5	415	2 AB0573	probable membrane
24	43	35.5	487	2 T06040	hypothetical prote
25	43	35.5	647	2 S67651	hypothetical prote
26	43	35.5	774	2 A13372	malate dehydrogena
27	43	35.5	903	2 T26743	hypothetical prote
28	42.5	35.1	174	2 G84600	hypothetical prote
29	42.5	35.1	878	2 B84977	alanine-tRNA ligas

30	42	34.7	82	2 T31574	hypothetical prote
31	42	34.7	194	2 B75613	conserved hypotet
32	42	34.7	227	2 B97624	hypothetical prote
33	42	34.7	268	2 S51505	cytochrome-c oxida
34	42	34.7	362	2 D90276	alcohol dehydrogen
35	42	34.7	523	2 T26740	hypothetical prote
36	42	34.7	531	2 T11074	hypothetical prote
37	42	34.7	544	2 B95412	hypothetical prote
38	42	34.7	646	2 T42296	hypothetical prote
39	41.5	34.3	520	2 T04550	hypothetical prote
40	41.5	34.3	554	1 TVYPBF	large T antigen -
41	41	33.9	124	2 S47145	neutonal protein -
42	41	33.9	132	2 H59093	hypothetical prote
43	41	33.9	157	2 H82154	conserved hypotet
44	41	33.9	275	2 F84088	N-acetylmutamoyl-L
45	41	33.9	307	2 AB2459	hypothetical prote

ALIGNMENTS

RESULT 1

H71031 probable DNA-directed RNA polymerase subunit A' - Pyrococcus horikoshii

C/Species: Pyrococcus horikoshii

C/Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #ext_change 21-Jul-2000

C/Accession: H71031

R/Kawabuchi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A/Reference number: A71000; PMID:98344137; PMID:9679194

A/Accession: H71031

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-907 <RAW>

A/Cross-reference: GB:AP000006; NID:G3236133; PIDN:BA030656.1; PID:G3257973

A/Experimental source: strain OT3

A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C/Genetics:

A/Gene: PH1545

C/Superfamily: Halobacterium DNA-directed RNA polymerase chain A

Query Match 40.5%; Score 49; DB 2; Length 907;

Best Local Similarity 52.2%; Pred. No. 18;

Matches 12; Conservative 4; Mismatches 3; Indels 4; Gaps 2;

QY 2 FHHT--FRGIVHG--KTIHRLV 20

Db 74 FGHTLARPVIHVGFAKTIHRL 96

RESULT 2

B75182 DNA-directed RNA polymerase, chain A' (rpoal) PAB0424 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #ext_change 20-Jun-2000

C/Accession: B75182

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A/Reference number: A75001

A/Accession: B75182

A/Status: preliminary

A/Molecule type: DNA

A/Cross-reference: GB:A248284; GB:AU096836; NID:G5457730; PIDN:CA049537.1; PID:G545804

A/Experimental source: strain Orsay

C/Genetics:

A/Gene: rpoal; PAB0424

C/Superfamily: Halobacterium DNA-directed RNA polymerase chain A

Query Match 40.5%; Score 49; DB 2; Length 907;

Best Local Similarity 52.2%; Pred. No. 18;
Matches 12; Conservative 4; Mismatches 3; Indels 4; Gaps 2;

QY 2 FHHI--FRGIYVVG--KTHRLV 20
Db 74 FGHILARPVIHVGPAKTHRL 96

RESULT 3

AD0273

Probable integral membrane protein YPO2242 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C/Accession: AD0273

R/Parikh, J.; Wren, B.W.; Thomson, N.R.; Tibbitt, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

Nature 413:523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AD0273

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-351 <KUR>

A/Cross-references: GB:AL590842; PIDN:CA091048.1; PID:G15980242; GSPDB:GN00175

C/Genetics: YPO2242

Query Match 39.7%; Score 48; DB 2; Length 351;
Best Local Similarity 50.0%; Pred. No. 9.1;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 5 IFRCIYHVGKTHRLVLTG 22
Db 160 IFQGYHSGENIHLQVQG 177

RESULT 4

S64944

hypothetical protein YLR107w - yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein L2904

C/Species: Saccharomyces cerevisiae

C/Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 19-Apr-2002

C/Accession: S64944; S69394

R/Vernasselt, P.; Voet, M.; Volckaert, G.

submitted to the Protein Sequence Database, May 1996

A/Reference number: S64943

A/Accession: S64944

A/Molecule type: DNA

A/Residues: 1-404 <VER>

A/Cross-references: EMBL:273279; NID:G1360495; PIDN:CAA97672.1; PID:e245554; PID:G136048

C/Genetics:

A/Genes: SGD:REX3

A/Cross-references: SGD:S0004097

A/Map position: 12R

Query Match 39.7%; Score 48; DB 2; Length 404;
Best Local Similarity 87.5%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFHHIFRG 8
Db 212 FFHHIFRG 219

RESULT 5
S43613

transposase (insertion sequence IS31831) - Corynebacterium glutamicum

C/Species: Corynebacterium glutamicum

C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-1999

C/Accession: S43613

R/Vertes, A.A.; Inui, M.; Kobayashi, M.; Kurusu, Y.; Yukawa, H.

Mol. Microbiol. 11, 739-746, 1994

A/Title: Isolation and characterization of IS31831, a transposable element from Coryneb

A/Reference number: S43613; MUID:94254729; PMID:8196545

A/Accession: S43613

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-436 <VER>

A/Cross-references: EMBL:D17429; NID:G473697; PIDN:BA04250.1; PID:dl004769; PID:g790952

Query Match 39.7%; Score 48; DB 2; Length 436;
Best Local Similarity 35.3%; Pred. No. 12;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHIFRCIYHVGKTHRL 19
Db 417 HYLRLIHSGQLVHKI 433

RESULT 6

G84634

Probable prolylcarboxypeptidase [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C/Accession: G84634

R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Noflat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: G84634

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-476 <STO>

A/Cross-references: GB:AE002093; NID:G6598559; PIDN:AAF18628.1; GSPDB:GN00139

C/Genetics:

A/Genes: At2g24280

A/Map position: 2

A/Superfamily: Caenorhabditis elegans ZK688.6 protein

Query Match 38.8%; Score 47; DB 2; Length 476;
Best Local Similarity 42.9%; Pred. No. 18;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FFHHIFRCIYHVGKTHRLV 21
Db 410 YFHOIFRVLNLISSIVALVT 430

RESULT 7

AC1208

GTP-binding elongation factor homolog [mo1067 [imported]] - Listeria monocytogenes (strain

C/Species: Listeria monocytogenes

C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C/Accession: AC1208

R/Glaser, P.; Frangoul, L.; Buchliesser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker,

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;

Science 294, 849-852, 2001

A/Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maicouram, A.; Mat

Ok, C.; Schlueter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A/Title: Comparative genomics of Listeria species.

A/Reference number: AB1077; MUID:21537279; PMID:11679669

A/Accession: AC1208

A/Status: preliminary

A/Molecule type: DNA

A;Residues: 1-612 <GUA>
A;Cross-references: GB:NC_003210; PIDN:CAC99145.1; PID:g16410469; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lml067
C;Superfamily: GTP-binding membrane protein lepa; translation elongation factor Tu homol

Query Match 38.8%; Score 47; DB 2; Length 612;
Best Local Similarity 58.3%; Pred. No. 24;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 IFRGIVHVGKTI 16
DB 230 VFRGTMHVGTIV 241

RESULT 8
AF1564
GTP-binding elongation factor homolog lml055 [imported] - listeria innocua (strain Clif
C;Species: listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1564
R;Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
D.; Dominguez Berni, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsih, H.
Science 294, 849-852, 2001
A;Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madueno, E.; Maltournam, A.; Ma
ok, C.; Schluerer, T.; Simoes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A;Title: Comparative genomics of listeria species
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1564
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-612 <GUA>
A;Cross-references: GB:AL592022; PIDN:CAC96286.1; PID:g16413514; GSPDB:GN00178
A;Experimental source: strain Clif11262
C;Genetics:
A;Gene: lml055
C;Superfamily: GTP-binding membrane protein lepa; translation elongation factor Tu homol

Query Match 38.8%; Score 47; DB 2; Length 612;
Best Local Similarity 58.3%; Pred. No. 24;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 IFRGIVHVGKTI 16
DB 230 VFRGTMHVGTIV 241

RESULT 9
S25564
DNA-directed RNA polymerase (EC 2.7.7.6) chain A' - Thermococcus celer
N;Alternate names: DNA-dependent RNA polymerase chain A'
C;Species: Thermococcus celer
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S25564; S27299
R;Klenk, H.P.; Schwass, V.; Lottspeich, F.; Zillig, W.
Nucleic Acids Res. 20, 4659, 1992
A;Title: Nucleotide sequence of the genes encoding the three largest subunits of the DNA
A;Reference number: S25362; MUID:93027167; PMID:11408768
A;Accession: S25564
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-905 <KLE>
A;Cross-references: EMBL:X67313; NID:g48138; PIDN:CAA47723.1; PID:g48141
A;Experimental source: DSM 2476
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1992
submitted to the Protein Sequence Database, February 1993
A;Reference number: S27298
A;Accession: S27299
A;Molecule type: protein
A;Residues: 2-18 <LOT>

C;Genetics:
A;Gene: rpoA1
C;Superfamily: Halobacterium DNA-directed RNA polymerase chain A
C;Keywords: nucleotidyltransferase; transcription
F;2-905/Product: DNA-directed RNA polymerase chain A' #status experimental <MAT>

Query Match 38.8%; Score 47; DB 1; Length 905;
Best Local Similarity 47.8%; Pred. No. 38;
Matches 11; Conservative 5; Mismatches 3; Indels 4; Gaps 2;

QY 2 FHHI--FRGIVHVG--KTIHRLV 20
DB 74 FGHVELARPVHVGAKTIHRLV 96

RESULT 10
C82309
probable inulinase-type protease VC0554 [imported] - vibrio cholerae (strain N16961 s
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: C82309
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; S
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: AB2035; MUID:20406833; PMID:10952301
A;Accession: C82309
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-952 <HEI>
A;Cross-references: GB:AE004141; GB:AE003852; NID:g9654976; PIDN:AAF93722.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0554
A;Map position: 1

Query Match 38.8%; Score 47; DB 2; Length 952;
Best Local Similarity 52.2%; Pred. No. 40;
Matches 12; Conservative 3; Mismatches 6; Indels 2; Gaps 2;

QY 1 FFRH-IFRGIVHVGKTIH-RLVT 21
DB 97 FFEHMFQGSKHVGDQHFRLIT 119

RESULT 11
B69354
ferritin homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Aug-1999
C;Accession: B69354
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Ueberback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.]
Smith, H.O.; Woes, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: B69354
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-173 <KLB>
A;Cross-references: GB:AE001047; GB:AE000782; NID:g2689370; PIDN:AAB90406.1; PID:g264977.
C;Superfamily: Ferritin

Query Match 38.0%; Score 46; DB 2; Length 173;
Best Local Similarity 47.4%; Pred. No. 8.5;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 FHHIFRGIVHVGKTIHRLV 20
DB 97 FFEHMFQGSKHVGDQHFRLIT 119

Db 89 FEHVEHEVNTKRIHELV 107

RESULT 12

T52621

mitogen-activated protein kinase MAP3K [Imported] - Arabidopsis thaliana (fragment)
C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 08-Dec-2000

C/Accession: T52621

R/Jouanvic, S.; Hamal, A.; Lepince, A.S.; Tregear, J.W.; Kreis, M.; Henry, Y.

Gene 229, 11-81, 1999

A/Title: Characterisation of novel plant genes encoding MEKK/STB1 and RAF-related proce
A/Reference number: Z24447; MUID:99196996; PMID:10095117

A/Accession: T52621

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-372 <JOU>

A/Cross-references: EMBL:Y14316; NID:G2315152; PIDN:CAA74696.1; PID:G2315153

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match

Best Local Similarity 38.0%; Score 46; DB 2; Length 372;
Matches 10; Conservative 4; Mismatches 8; Indels 2; Gaps 1;

Qy 1 FFHHIFRGI--VHVGKTIHRLVVG 22

Db 108 FTRHILSGIAYHNKKTVHRDIKG 131

RESULT 13

T36329

probable GTP cyclohydrolase I - Streptomyces coelicolor
C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C/Accession: T36329

R/Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999

A/Reference number: Z21575

A/Accession: T36329

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-201 <OLI>
A/Cross-references: EMBL:AL049841; PIDN:CAM42756.1; GSPDB:GN00070; SCOPDB:SCB9.10C

A/Experimental source: strain AJ(2)

C/Genetics:

A/Gene: folE; SCOPDB:SCB9.10C

C/Superfamily: GTP cyclohydrolase I

Query Match

Best Local Similarity 37.2%; Score 45; DB 2; Length 201;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

Qy 3 HHIFRGIHVHVG 13

Db 92 HHLVPRGVAVHG 104

RESULT 14

JC4742

transposase - Corynebacterium glutamicum

C/Species: Corynebacterium glutamicum

C/Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 17-Mar-1999

C/Accession: JC4742

R/Correia, A.; Pisabarro, A.; Castro, J.M.; Martin, J.F.

Gene 170, 91-94, 1996

A/Title: Cloning and characterization of an IS-like element present in the genome of Bre

A/Reference number: JC4742; MUID:96200862; PMID:8621097

A/Accession: JC4742

A/Molecule type: DNA

A/Residues: 1-436 <COR>

A/Cross-references: EMBL:Z66534

A/Experimental source: ATCC 13869

A/Note: The authors translated the initiation codon TGA for residue 1 as Val

A/Note: The authors translated the codon ATT for residue 125 as Tyr

A/Note: the source is designated as Brevibacterium lactofermentum

A/Genetics:

A/Gene: GTG

F/388-415/Domain: DNA binding #status predicted <DNA>

F/405-415/Region: helix-turn-helix

Query Match

Best Local Similarity 36.4%; Score 44; DB 2; Length 436;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HHIFRGIHVHVGKTIHRL 19

Db 417 HYLRLCHSGQLTKHI 433

RESULT 15

C75553

hypothetical protein - Deinococcus radiodurans (strain R1)

C/Species: Deinococcus radiodurans

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000

C/Accession: C75553

R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A/Reference number: A75250; MUID:20036896; PMID:10567266

A/Accession: C75553

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-513 <WHI>

A/Cross-references: GB:AE001878; GB:AE000513; NID:G6457820; PIDN:AR09750.1; PID:G645782

A/Experimental source: strain R1

C/Genetics:

A/Map position: 1

C/Superfamily: Deinococcus radiodurans hypothetical protein DR0157

Query Match

Best Local Similarity 36.4%; Score 44; DB 2; Length 513;
Matches 11; Conservative 1; Mismatches 4; Indels 8; Gaps 1;

Qy 2 FFHHIFRGIHV-----GKTIH 17

Db 103 FHLRFGVHVAHMLAPGKDIH 126

Search completed: May 23, 2003, 08:57:17

Job time : 17 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2003, 08:52:09 ; Search time 7.5 Seconds
(without alignments)
121.664 Million cell updates/sec

Title: US-09-929-788-2

Perfect score: 121
Sequence: 1 FFHHIFRGIVHGKTHRLVYG 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	39.7	351	1 RNFD_YERPE	O82d2 yersinia pe
2	47	38.8	905	1 RPA1_THECE	P31813 thermococ
3	46	38.0	501	1 NR52_CHICK	O42101 gallus gall
4	45	37.2	201	1 GCH1_STRCO	O92813 streptomyc
5	44	36.4	666	1 PD14_MOUSE	O92183 mus musculu
6	44	36.4	917	1 VGLB_HSVB2	P12641 bovine herp
7	43	35.5	192	1 EM3R_BACME	P33506 bacillus me
8	43	35.5	219	1 CLN1_HUMAN	O9bux9 homo sapien
9	43	35.5	219	1 CLN1_MOUSE	O9j357 mus musculu
10	43	35.5	322	1 GDC_RAT	P16261 ratu
11	43	35.5	330	1 GDC_BOVIN	P57483 buchnera ap
12	42.5	35.1	878	1 SYA_BUCAL	O01888 bos taurus
13	42	34.7	663	1 PD15_HUMAN	O9um67 homo sapien
14	42	34.7	666	1 PD14_RAT	O88807 ratu
15	41.5	34.3	587	1 TALA_BPDV	P13834 budgerigar
16	41	33.9	124	1 NEUR_FELCA	P41737 felis silve
17	41	33.9	213	1 GCH1_OSTOS	O61573 osterlagia
18	41	33.9	311	1 PYRD_STRPN	O92950 streptococ
19	41	33.9	332	1 GDC_HUMAN	P16260 homo sapien
20	41	33.9	398	1 HEMI_MERTM	P42809 methanobact
21	41	33.9	402	1 HEMI_MERTH	P42703 methanobact
22	41	33.9	410	1 COTR_CAVPO	P22333 cavia porce
23	41	33.9	412	1 DAD3_RHILIO	O981x2 thizobium 1
24	41	33.9	496	1 HUTH_THECAC	O9h116 thermoplasm
25	41	33.9	3829	1 SACS_HUMAN	O9nz14 homo sapien
26	41	33.9	3830	1 SACS_MOUSE	O9j1c8 mus musculu
27	40.5	33.5	533	1 N1FD_CLOPA	P06273 clostridium
28	40	33.1	202	1 GCH1_MYCTU	O06273 mycobacteri
29	40	33.1	220	1 GLT2_ARATH	O91mc9 arabidopsis
30	40	33.1	223	1 GCH1_CAEBL	O91980 caenorhabdi
31	40	33.1	234	1 CYPR_CALOVI	P86517 calliphora
32	40	33.1	257	1 NUDC_ECO57	O9X6x7 escherichia
33	40	33.1	257	1 NUDC_ECOLI	P32664 escherichia

34	40	33.1	257	1 NUDC_SALTI	O82328 salmonella
35	40	33.1	257	1 NUDC_SALTY	O91915 salmonella
36	40	33.1	260	1 YQ08_CAEBL	O09297 caenorhabdi
37	40	33.1	297	1 CDC2_MOUSE	P11440 mus musculu
38	40	33.1	329	1 Y493_MYCTU	O11458 mycobacteri
39	40	33.1	342	1 TRM1_THEVO	O97az2 thermoplasm
40	40	33.1	518	1 GSH1_BUCAL	P57485 buchnera ap
41	40	33.1	539	1 FUT2_ARATH	O81053 arabidopsis
42	40	33.1	655	1 Y068_YEAST	P47139 saccharomyc
43	40	33.1	810	1 NEB1_HUMAN	O92833 homo sapien
44	40	33.1	1080	1 M115_CAEBL	O23356 caenorhabdi
45	39.5	32.6	60	1 RU30_STRCO	P46789 streptomyc

ALIGNMENTS

```

RESULT 1
RNFD_YERPE
ID RNFD_YERPE STANDARD; PRT; 351 AA.
AC O82ED2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfd.
GN RNFD OR YP02242.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tlthball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds W., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,
CC rnfd, rnfe and rnfg (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE NORB/RNFD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A414151; CAC91048.1; -
DR InterPro; IPR004338; NOR2_Rnfd_Rnfe.
DR Pfam; PF03116; NOR2_Rnfd_Rnfe; 1.
KW Electron transport; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 21
FT TRANSMEM 93
FT TRANSMEM 115
FT TRANSMEM 122
FT TRANSMEM 144
FT TRANSMEM 210
FT TRANSMEM 232
FT TRANSMEM 237
FT TRANSMEM 259
FT TRANSMEM 269
FT TRANSMEM 291
FT TRANSMEM 315
SQ SEQUENCE 351 AA; 38112 MW; A35EAE7524F605F8 CRC64;
Query Match 39.7%; Score 48; DB 1; Length 351;

```



```

RX MEDLINE=21996410; PubMed=12000953;
RA Bertline S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Houchard S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC (erythro-1,2,3-ethylidioxypyrrolidinohydroxyphenyl) triphosphate.
CC -1- SUBUNIT: HOMOPOLYMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL049841; CAB42756.1; -.
DR HSSP; P27511; 1ABR.
DR InterPro; IPR001474; GTP_cyclohydrol.
DR Pfam; PF01227; GTP_cyclohydrol; 1.
DR ProDom; PD003330; GTP_cyclohydrol; 1.
DR TIGRFAMs; TIGR00063; fole; 1.
DR PROSITE; PS00859; GTP_CYCLOHYDROL_1; 1.
DR PROSITE; PS00860; GTP_CYCLOHYDROL_2; 1.
DR One-carbon metabolism; Hydrolyase; Complete proteome.
FT DISUFID 90 163
SQ SEQUENCE 201 AA; 22415 MW; 4CC8FE676687B6B CRC64;

Query Match 37.2%; Score 45; DB 1; Length 201;
Best Local Similarity 61.5%; Pred. No. 5.9;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 3 HHT--FRGIVAVG 13
DB 92 HHLVFRGVAHV 104

RESULT 5
PD14_MOUSE STANDARD; PRT; 666 AA.
AC Q92183;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-arginine deiminase type IV (EC 3.5.3.15) (Peptidylarginine
DE deiminase IV).
GN PAD14 OR PD14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epidermis;
RX MEDLINE=99192810; PubMed=10092850;
RA Rusd A.A., Ikejiri Y., Ono H., Yonekawa T., Shiraiwa M., Kawada A.,
RA Takahara H.;
RT "Molecular cloning of cDNAs of mouse peptidylarginine deiminase type
RT I, type III and type IV, and the expression pattern of type I in
RT mouse.";
RL Eur. J. Biochem. 259:660-669(1999).
CC -1- FUNCTION: CATALYZES THE DEIMINATION OF ARGININE RESIDUES OF

```

```

CC PROTEINS.
CC -1- CATALYTIC ACTIVITY: Protein L-arginine + H(2)O = protein L-
CC citrulline + NH(3).
CC -1- COFACTOR: REQUIRES CALCIUM IONS.
CC -1- SIMILARITY: SOME SIMILARITY WITH ARGINASE AND TO ARGININE
CC DEIMINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB013850; BAA3246.1; -.
DR MGD; MG1:1338898; Pd14.
DR InterPro; IPR004303; Prot_arg_deim.
DR Pfam; PF03068; PAD; 1.
DR TIGRFAMs; TIGR00063; fole; 1.
DR PROSITE; PS00859; GTP_CYCLOHYDROL_1; 1.
DR PROSITE; PS00860; GTP_CYCLOHYDROL_2; 1.
DR One-carbon metabolism; Hydrolyase; Complete proteome.
FT DISUFID 90 163
SQ SEQUENCE 666 AA; 74476 MW; 70FAE4E7E232D34A CRC64;

Query Match 36.4%; Score 44; DB 1; Length 666;
Best Local Similarity 33.3%; Pred. No. 29;
Matches 6; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FPHHFRGIVHVKTIHR 18
DB 637 YYYHNGEVRHCGTNVRR 654

RESULT 6
VGBL_HSVB2 STANDARD; PRT; 917 AA.
AC P12641;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein B-1 precursor.
DE Bovine herpesvirus type 2 (strain BWV) (Bovine mammillitis virus).
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89306231; PubMed=2841793;
RA Hamerschmidt W., Contraths F., Mankertz J., Pauli G., Ludwig H.,
RA Buhk H.-J.;
RT "Conservation of a gene cluster including glycoprotein B in bovine
RT herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL Virology 165:388-405(1988).
RN [2]
RP SEQUENCE OF 1-200 FROM N.A.
RX MEDLINE=88306232; PubMed=2457278;
RA Hamerschmidt W., Contraths F., Mankertz J., Buhk H.-J., Pauli G.,
RA Ludwig H.;
RT "Common epitopes of glycoprotein B map within the major DNA-binding
RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT virus type 1 (HSV-1).";
RL Virology 165:406-418(1988).
CC -1- FUNCTION: GBI IS A 130 kDa GLYCOPROTEIN WHICH IS NECESSARY FOR THE
CC PENETRATION OF THE VIRUS INTO THE HOST CELL AND THE INDUCTION OF A
CC SYNCTIAL PHENOTYPE.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUS GLYCOPROTEIN B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL, M21628; AAA46053.1; -
DR EMBL, M21632; AAA46052.1; -
DR PIR, C29242; VGBEBH.
DR InterPro; IPR000234; Glycoprot B.
DR Pfam; PF00606; Glycoprotein_B_1.
DR ProDom; PD006093; Glycoprot_B_1.
KM Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 917 GLYCOPROTEIN B-1.
FT TRANSMEM 578 594 POTENTIAL.
FT TRANSMEM 770 786 POTENTIAL.
FT TRANSMEM 795 811 POTENTIAL.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 917 AA; 101882 MW; 1B96CBF50DB4D3F3 CRC64;

Query Match 36.4%; Score 44; DB 1; Length 917;
Best Local Similarity 40.0%; Pred. No. 41;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 HHIFRGIVGKTHRLVLT 22
Db 754 HGFEGLDGVRGVRVVLG 773

RESULT 7
BM3R_BACME STANDARD; PRT; 192 AA.
ID BM3R_BACME
AC P43506;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transcriptional repressor Bm3R1.
GN Bm3R1.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1404;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92184811; PubMed=1544926;
RA Shaw G.C., Fulco A.J.;
RT "Barbiturate-mediated regulation of expression of the cytochrome P450BM-3 gene of Bacillus megaterium by Bm3R1 protein."
RL J. Biol. Chem. 267:5515-5526(1992).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=89291834; PubMed=2544578;
RA Ruettinger R.T., Wen L.P., Fulco A.J.;
RT "Coding nucleotide, 5' regulatory, and deduced amino acid sequences of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450 reductase from Bacillus megaterium."
RL J. Biol. Chem. 264:10987-10995(1989).
RN (3)
RP CHARACTERIZATION.
RX MEDLINE=93155125; PubMed=8428974;
RA Shaw G.C., Fulco A.J.;
RT "Inhibition by barbiturates of the binding of Bm3R1 repressor to its operator site on the barbiturate-inducible cytochrome P450BM-3 gene of Bacillus megaterium."
RL J. Biol. Chem. 268:2997-3004(1993).
CC -1- FUNCTION: NEGATIVELY CONTROLS THE EXPRESSION OF THE CYTOCHROME P450BM-3 GENE AT THE TRANSCRIPTIONAL LEVEL.
CC -1- SIMILARITY: BELONGS TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL REGULATORS.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, S87512; AAB21757.1; -
DR EMBL, J04832; AAA87601.1; -
DR InterPro; IPR001647; HTH_Tetr.
DR Pfam; PF00440; tetr. 1.
DR PRINTS; PR00455; HTHTEPR.
DR PROSITE; PS01081; HTH TETR FAMILY: 1.
KM Transcription regulation; Repressor; DNA-binding.
FT DNA_BIND 28 47 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 192 AA; 21886 MW; 766AC6D34944748 CRC64;

Query Match 35.5%; Score 43; DB 1; Length 192;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FHIFRGIV 10
Db 84 FHIFRGIV 92

RESULT 8
CALN1_HUMAN STANDARD; PRT; 219 AA.
ID CALN1_HUMAN
AC Q9BXU9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calneuron 1.
GN CALN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21185206; PubMed=11286509;
RA Wu Y.-O., Lin X., Liu C.-M., Jamrich M., Shaffer L.G.;
RT "Identification of a human brain-specific gene, calneuron 1, a new member of the calmodulin superfamily."
RL Mol. Genet. Metab. 72:343-350(2001).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strassberg R.;
RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May play a role in the physiology of neurons and is potentially important in memory and learning.
CC -1- TISSUE SPECIFICITY: Brain-specific.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF282250; AAK15155.1; -
DR EMBL, BC020200; AAH20200.1; -
DR HSSP; P02593; 1FM4.
DR Genew; HGNC:13248; CALN1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR ProDom; PD000012; EF-hand; 1.

```

```

DR SMART; SM00054; EFh; 2.
KW PROSITE; PS00018; EF-HAND; 2.
KM Calcium-binding; Repeat.
FT CA BIND 49 60 EF-HAND 1 (POTENTIAL).
FT CA BIND 85 96 EF-HAND 2 (POTENTIAL).
SQ SEQUENCE 219 AA; 24837 MW; EAD2BE2090F5CA0F CRC64;

Query Match 35.5%; Score 43; DB 1; Length 219;
Best Local Similarity 35.3%; Pred. No. 13;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 FHIFRGIVHVGKTIHR 18
Db 3 FHHVTAGLVKGNVYLR 19

RESULT 9
CLN1_MOUSE STANDARD; PRT; 219 AA.
AC 09JUG7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calneuron 1.
GN CALN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=21185206; PubMed=11286509;
RA Wu Y.-Q., Lin X., Liu C.-M., Jamrich M., Shaffer L.G.;
RT "Identification of a human brain-specific gene, calneuron 1, a new
RT member of the calmodulin superfamily."
RL Mol. Genet. Metab. 72:343-350(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Oseida N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
RT made by oligo-capping method."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May play a role in the physiology of neurons and is
CC potentially important in memory and learning.
CC -1- TISSUE SPECIFICITY: Brain-specific. High expression in the
CC cerebellum, hippocampus, and cortex.
CC -1- DEVELOPMENTAL STAGE: Shows little prenatal expression, with
CC highest expression at postnatal day 21.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF282251; AAK15156.1; -
DR EMBL; AB041539; BA95024.1; -
DR HSSP; P02593; 1PW4.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR ProDom; PD000012; EF-hand; 1.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF-HAND; 2.
KW Calcium-binding; Repeat.
FT CA BIND 49 60 EF-HAND (POTENTIAL).
FT CA BIND 85 96 EF-HAND (POTENTIAL).
SQ SEQUENCE 219 AA; 24823 MW; F43DA0D17FEB250F CRC64;

```

```

Query Match 35.5%; Score 43; DB 1; Length 219;
Best Local Similarity 35.3%; Pred. No. 13;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 FHIFRGIVHVGKTIHR 18
Db 3 FHHVTAGLVKGNVYLR 19

RESULT 10
GDC_RAT STANDARD; PRT; 322 AA.
ID GDC_RAT
AC P16261;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Grave's disease carrier protein (GDC) (Mitochondrial solute carrier
DE protein homolog) (Fragment).
GN SLC25A16 OR GDA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=90114217; PubMed=2575220;
RA Zaritelli R., Oates E.L., McBride O.W., Lerman M.I., Chan J.Y.,
RA Santisteban P., Ursini M.V., Notkins A.L., Kohn L.D.;
RT "Sequence and chromosomal assignment of a novel cDNA identified by
RT immunoscreening of a thyroid expression library: similarity to a
RT family of mitochondrial solute carrier proteins."
RL Mol. Endocrinol. 3:1498-1508(1989).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (Potential).
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M32973; AAA41639.1; -
DR InterPro; IPR001993; Mitoch carrier.
DR Pfam; PF00153; mito carr; 3.
DR PROSITE; PS00215; MITOCH_CARRIER; 3.
KW Mitochondrion; Transmembrane; Transport; Repeat.
FT REPEAT 1 123 1.
FT REPEAT 124 222 2.
FT REPEAT 223 >322 3.
FT NON TER 322 322
SQ SEQUENCE 322 AA; 35056 MW; F78CBA36CA9DC9A CRC64;

Query Match 35.5%; Score 43; DB 1; Length 322;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 HIFRGIVHVGKTIHR 17
Db 169 HHTYSGIHTAFKTIY 182

RESULT 11
GDC_BOVIN STANDARD; PRT; 330 AA.
ID GDC_BOVIN
AC 001889;
DT 01-APR-1993 (Rel. 25, Created)

```

```

DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Grave's disease carrier protein (GDC) (Mitochondrial solute carrier
DE protein homolog).
GN SLC25A16 OR GDA OR GDC.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=93091248; PubMed=1457817;
RA Fiermonte G., Runswick M.J., Walker J.E., Palmieri F.;
RT "Sequence and pattern of expression of a bovine homologue of a human
RT mitochondrial transport protein associated with Grave's disease.";
RL DNA Seq. 3:71-78(1992).
CC
CC -1- FUNCTION: Required for the accumulation of coenzyme A in the
CC mitochondrial matrix (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (By similarity).
CC -1- TISSUE SPECIFICITY: MOSTLY IN THYROID, LIVER, LUNG, KIDNEY AND
CC TO A LESSER EXTENT IN HEART AND SKELETAL MUSCLE.
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X66035; CAA46834.1; -.
DR PIR: S26596; S26596.
DR InterPro: IPR002067; Mlt_carrier.
DR InterPro: IPR001993; Mitoch_carrier.
DR Pfam: PF00153; mltc_carr; 3.
DR PRINTS: PR00926; MITOCH_CARRIER.
DR PROSITE: PS00215; MITOCH_CARRIER.
DR Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
KW TRANSMEM 132 152 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT REPEAT 232 252 POTENTIAL.
FT REPEAT 1 121 1.
FT REPEAT 122 217 2.
FT REPEAT 218 330 3.
SQ SEQUENCE 330 AA; 36085 MW; 4C614701D2B8DEA9 CRC64;

Query Match 35.5%; Score 43; DB 1; Length 330;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

RN
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC
CC -1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA (Ala) = AMP +
CC diphosphate + L-alanyl-tRNA (Ala).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AP001119; BAB13106.1; -.
DR InterPro: IPR002106; AAcRNA_ligaseII.
DR InterPro: IPR003156; DHHA1.
DR InterPro: IPR002318; tRNA-synt_2c.
DR Pfam: PF01411; tRNA-synt_2c; 1.
DR Pfam: PF02272; DHHA1; 1.
DR PRINTS: PR00980; TRNASYNTHALA.
DR TIGRfam: TIGR00344; alas; 1.
DR PROSITE: PS50860; AA_TRNA_LIGASE_II_ALA_1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 878 AA; 101402 MW; 005321303125D165 CRC64;

Query Match 35.1%; Score 42.5; DB 1; Length 878;
Best Local Similarity 50.0%; Pred. No. 67;
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017919; BAA84542.1; -
DR GenBank; HGNC:18342; PAD15.
DR MIM; 605347; -
DR InterPro; IPR004303; Prot_arg_deim.
DR Pfam; PF03068; PAD; 1.
KM Hydrolase; Calcium-binding; Multigene family.
FT CA BIND EF-HAND (POTENTIAL).
FT CA BIND 505 516
SQ SEQUENCE 663 AA; 74095 MW; 45FA31F4ACFD3CDB CRC64;

Query Match 34.7%; Score 42; DB 1; Length 663;
Best Local Similarity 44.4%; Pred. No. 60;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 FFHHIFRGIVHVKTIHR 18
Db 634 FTYHIRHGEVHCNTVRR 651

RESULT 14
PD14_RAT STANDARD; PRT; 666 AA.
ID PD14_RAT
AC O88807; 035117;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-arginine deiminase type IV (EC 3.5.3.15) (Peptidylarginine
DE deiminase IV) (PAD-R4) (Peptidylarginine deiminase type alpha).
GN PAD14 OR PD14.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409324; PubMed=9738944;
RA Ishigami A., Kuramoto M., Yamada M., Matanabe K., Senshu T.;
RT "Molecular cloning of two novel types of peptidylarginine deiminase
RT cDNAs from retinoic acid-treated culture of a newborn rat
RT keratinocyte cell line."
RL FEBS Lett. 433:113-118(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Epidermis;
RX MEDLINE=98342160; PubMed=9675292;
RA Yamakoshi A., Ono H., Nishijo T., Shiraiwa M., Takahara H.;
RT "Cloning of cDNA encoding a novel isoform (type IV) of
RT peptidylarginine deiminase from rat epidermis."
RL Biochim. Biophys. Acta 1386:227-232(1998).
CC -1- FUNCTION: CATALYZES THE DEIMINATION OF ARGININE RESIDUES OF
CC PROTEINS.
CC -1- CATALYTIC ACTIVITY: Protein L-arginine + H(2)O = protein L-
CC citrulline + NH(3).
CC -1- COFACTOR: REQUIRES CALCIUM IONS.
CC -1- TISSUE SPECIFICITY: EPIDERMIS.
CC -1- SIMILARITY: SOME SIMILARITY WITH ARGINASE AND TO ARGININE
CC DEIMINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AB010999; BAA32100.1; -
DR EMBL; AB008803; BAA23523.1; -
DR InterPro; IPR004303; Prot_arg_deim.
DR Pfam; PF03068; PAD; 1.
KM Hydrolase; Calcium-binding; Multigene family.
FT CA BIND EF-HAND (POTENTIAL).
FT CONFLICT 655 655 K -> E (IN REF. 2).
SQ SEQUENCE 666 AA; 74467 MW; 70F23BB3B04C813B CRC64;

Query Match 34.7%; Score 42; DB 1; Length 666;
Best Local Similarity 33.3%; Pred. No. 60;
Matches 6; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 FFHHIFRGIVHVKTIHR 18
Db 637 YSYHMYHGEVHCNTVRR 654

RESULT 15
TALA_BFDV STANDARD; PRT; 587 AA.
ID TALA_BFDV
AC P13854;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Large T antigen.
DE Bacteriophage phi801 (BFDV).
OS Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88265888; PubMed=2838972;
RA Rott O., Kroege M., Mueller H., Hobom G.;
RT "The genome of bacteriophage phi801 (BFDV), an avian
RT polyomavirus."
RL Virology 165:74-86(1988).
RN [2]
RP REVISIONS.
RA Stoll R., Dong L., Kouwenhoven B., Hobom G., Mueller H.;
RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20775; AAB59760.1; -
DR PIR; D29194; TVVPBF.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_2; 1. FALSE_NEG.
DR PROSITE; PS50076; DnaJ_2; 1.
KM Early protein; Nuclear protein; ATP-binding; DNA-binding;
KM DNA replication; Zinc-finger.
FT DOMAIN 6 82 J-DOMAIN.
FT ZN BIND 258 282 C2H2-TYPE (ATYPICAL) (BY SIMILARITY).
FT NP BIND 386 393 ATP (POTENTIAL).
SQ SEQUENCE 587 AA; 67132 MW; 73E5D71DA92EB873 CRC64;

Query Match 34.3%; Score 41.5; DB 1; Length 587;
Best Local Similarity 44.0%; Pred. No. 63;
Matches 11; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

Qy 3 HHIFRGIVHVKTIHR 22
Db 381 YIFKGPVNTGKTVAAAILALCTG 405

Tue May 27 13:06:29 2003

us-09-929-788-2.rsp

Page 8

Search completed: May 23, 2003, 08:55:45
Job time : 9.5 secs

DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Moronecidin prepropeptide precursor.
 OS Morone chrysoptera (white bass).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percidae;
 OC Moronidae; Morone.
 NCBI_TaxID=46259;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lauth X.C., Shike H., Burns J.C., Westerman M.E., Ostland V.E.,
 RA Carlberg J.M., Van Olst J.C., Nizet V., Taylor S.W., Shimizu C.,
 RA Bulet P.;
 RT "Discovery and characterization of two isoforms of moronecidin, a
 RT novel antimicrobial peptide from hybrid striped bass.";
 RL J. Biol. Chem. 0:0-0(2002).
 DR EMBL: AF332621; AAL40409.1; -;
 DR EMBL: AF394243; AAL57318.1; -;
 KW Signal.
 FT SIGNAL.
 FT CHAIN 23 44 POTENTIAL.
 SQ SEQUENCE 79 AA; 9106 MW; B6F32481C4200EBB CRC64;

Query Match 97.5%; Score 118; DB 13; Length 79;
 Best Local Similarity 95.5%; Pred. No. 2,1e-11;
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFHHIFRGIVHVKTIHRLVTG 22
 DB 23 FFHHIFRGIVHVKTIHRLVTG 44

RESULT 3
 Q9REI4 PRELIMINARY; PRT; 28 AA.
 AC Q9REI4;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE Hypothetical 3.4 kDa protein.
 OS Acidiphilium symbioticum.
 OC Acidiphilium PAS3.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Acidiphilium.
 NCBI_TaxID=94005;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mahapatra N.R., Ghosh S., Deb C., Banerjee P.C.;
 RT "Cloning, sequencing and analysis of a cadmium and zinc resistance
 RT conferring plasmid segment from Acidiphilium symbioticum KM2.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ239066; CAB65723.1; -;
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 28 AA; 3379 MW; CA15E234000EC599 CRC64;

Query Match 41.3%; Score 50; DB 2; Length 28;
 Best Local Similarity 47.1%; Pred. No. 0.53;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 HHIRGIVHVKTIHRL 19
 DB 6 HRLQALFHIGTTFHRL 22

RESULT 4
 Q9Z486 PRELIMINARY; PRT; 698 AA.
 AC Q9Z486;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Alpha-glucuronidase.
 GN xyg.
 OS Aeromonas punctata (Aeromonas caviae).
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 NCBI_TaxID=648;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ME-1;
 RA Kitagawa E., Suzuki T., Kawai K.;
 RT "Xylosidase gene(xysB) and alpha-glucuronidase gene(xyg) of Aeromonas
 RT caviae ME-1.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB022788; BAA74508.1; -;
 DR InterPro: IPR005154; Glyco_hydro. 67.
 DR Pfam: PF03648; Glyco_hydro. 67; 1-
 SQ SEQUENCE 698 AA; 78166 MW; 064958F833654252 CRC64;

Query Match 40.9%; Score 49.5; DB 2; Length 698;
 Best Local Similarity 47.1%; Pred. No. 21;
 Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

QY 1 FFHHIFRGIVHVKTI 16
 DB 623 FFHHIFRGIVHVKTI 639

RESULT 5
 Q9FW78 PRELIMINARY; PRT; 537 AA.
 AC Q9FW78;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Putative cytochrome P450.
 GN OSJNB002612.15.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriacoidae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Buehl C.R., Yuan Q., Moffat K.S., Hall J.N., Jenkins C.N., Burr P.C.,
 RA Hsiao J., Zisman V., Pai G., Bowman C.L., Fujii C.Y., Vankken S.B.,
 RA Bowman C.L., Craven B., Uteback T.R., Khalak H., Feldblum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNB002612 genomic sequence.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AC068924; AAG33500.1; -;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00067; p450.1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Hem; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 537 AA; 59970 MW; 760CF5A958C27383 CRC64;

Query Match 40.5%; Score 49; DB 10; Length 537;
 Best Local Similarity 53.3%; Pred. No. 19;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 FFHIFR-----GIVHVKTIHRLVT 21
 Db 167 FYGHFLREITVATGTFITGEVHMILT 194

RESULT 10

Q82ED2 PRELIMINARY; PRT; 351 AA.
 AC Q82ED2;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Putative integral membrane protein.
 GN YPO2242.
 GN Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OC NCBI_TaxID=632;
 RX [1]

SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,
 RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Baaham D., Bentley S.D., Brooks K., Cerdono-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Letwell T., Martin N., Holroyd S.C., Jagers K., Karlyshev A.V.,
 RA Leather S., Moulton J., Oyston P.C.F., Quail M., Rutherford B.G.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett K.G.,
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague.",
 RL Nature 413:523-527(2001).
 DR EMBL; AJ141451; CAC91048.1; -;
 DR InterPro; IPR004338; NCR2_RnFf_RnFf.
 DR Pfam; PF03116; NCR2_RnFf_RnFf.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 351 AA; 36112 MW; A35EAE7524F605F8 CRC64;

Query Match 39.7%; Score 48; DB 16; Length 351;
 Best Local Similarity 50.0%; Pred. No. 17;
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 5 IFRGIVHVKTIHRLVNG 22
 Db 160 IFTGYTHSGENIHQLQVNG 177

RESULT 11

Q12090 PRELIMINARY; PRT; 404 AA.
 AC Q12090;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Similar to S. POMER hypothetical protein C2267.04P (YLR107W).
 GN REX3 OR L8004.1 OR L2904 OR YLR107W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 OC NCBI_TaxID=4932;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Geisel C.;

Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

SEQUENCE FROM N.A.
 RC STRAIN=S288C;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gatungu S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,

RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
 RA Wilson R., Waterston R.,
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]

SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Waterston R.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]

SEQUENCE FROM N.A.
 RC STRAIN=FY23/RD005;
 RA Vernasselt P., Volckaert G.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]

SEQUENCE FROM N.A.
 RA WPS;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [6]

SEQUENCE OF 275-404 FROM N.A.

RC STRAIN=S288C;
 RX MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albertmann K., Andrie B., Ansoorge W.,
 RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
 RA Ertlan K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter P.,
 RA Louis E.J., Messenguy F., Mewes H.W., Miosga T., Mostl D.,
 RA Muller-Auer S., Nentwich B., Obermaier B., Pflavandl E., Pohl T.M.,
 RA Portetle D., Purcell B., Reckman S., Rieger M., Rinke M., Rose M.,
 RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
 RA Underwood A.P., Uristarazu L.A., Vandenbol M., Verasselt P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohnsels J.D., XII.,
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII.",
 RL Nature 387:0-0(0).

SEQUENCE OF 275-404 FROM N.A.

RC STRAIN=S288C;

RA Cherry J.M.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53876; AAB67549.1; -;

DR EMBL; X89514; CAA61685.1; -;

DR EMBL; Z73279; CAA97672.1; -;

DR EMBL; U53878; AAB67561.1; -;

DR EMBL; Z73280; CAA97675.1; -;

DR SGD; S0004097; REX3.

DR InterPro; IPR000520; Exonuclease.

DR Pfam; PF00929; Exonuclease; 1.

DR SMART; SM00479; EXOIII; 1.

SEQUENCE 404 AA; 45879 MW; FF6DF94ECA69A5F CRC64;

Query Match 39.7%; Score 48; DB 3; Length 404;
 Best Local Similarity 87.5%; Pred. No. 20;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FFHIFRG 8
 Db 212 FFHIFRG 219

RESULT 12

Q45144 PRELIMINARY; PRT; 436 AA.

AC Q45144;
 AC Q45144;

DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE DNA, transposable element IS11831 (Transposase).

GN TPN.

OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;

OC Corynebacterium.

OC NCBI_TaxID=1718;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC31831;
 RA MEDLINE=94254729; PubMed=8196545;
 RX Vertes A., Inui M., Kobayashi M., Kurusu Y., Yukawa H.;
 RT "Isolation and characterization of ISJ1831 a transposable element from
 RL Corynebacterium glutamicum".
 RN Mol. Microbiol. 11:739-746(1994).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CN6;
 RC
 RA Ninomiya K., Garbe T.R., Inui M., Yukawa H.;
 RT "Survival of High-dose Cytidine Associated with Transposition in
 RL Corynebacterium glutamicum".
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D17429; BAA04250.1; -;
 DR EMBL; AF435074; ALJ1543.1; -;
 DR EMBL; AF435074; ALJ1536.1; -;
 DR InterPro: IPR002560; Transposase_12.
 DR Pfam; PFO1610; Transposase_12; 1.
 DR DNA-binding.
 QO SEQUENCE 436 AA; 49576 MW; 116D3A78DCADB5A CRC64;

Query Match	39.7%	Score 48;	DB 2;	Length 436;
Best Local Similarity	35.3%;	Pred. No. 22;		
Matches	6;	Conservative	7;	Mismatches 4;
				Indels 0;
				Gaps 0.

RESULT 13	
090XG6	PRELIMINARY; PRT; 1775 AA.
ID 090XG6	
AC 090XG6	
DT 01-DEC-2001	(TREMBLrel. 19, Created)
DT 01-DEC-2001	(TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002	(TREMBLrel. 21, Last annotation update)
GN Myosin IIIA.	
GN MIO3A.	
OS Brachydanio rerio (Zebrafish) (Zebra danio).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;	
OC Cyprinidae; Danio.	
OX NCBI_TaxId=7955;	
RN [1]	
RA SEQUENCE FROM N.A.	
RP Solhberg L., Burnside B.;	
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.	
DR EMBL: AF84863; AAL1513.1; -	
DR InterPro: IPR000719; Euk_pkinase.	
DR InterPro: IPR000048; IQ_region.	
DR InterPro: IPR001609; myosin_head.	
DR InterPro: IPR001245; Tyr_pkinase.	
DR Pfam: PF00612; IQ: 9.	
DR Pfam: PF00063; myosin_head; 1.	
DR Pfam: PF00063; pkinase; 1.	
DR ProDom: PD000001; Euk_pkinase; 1.	
DR ProDom: PD000355; myosin_head; 1.	
DR SMART: SM00219; TyrKc; 1.	
DR PROSITE: PS50096; IQ: 6.	
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.	
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.	
DR ATP-binding; Transferrase.	
Q1 SEQUENCE 1775 AA; 202415 MW; 0AD184FEFA802629 CRC64;	

Query Match	39.7%	Score 48;	DB 13;	Length 1775;
Best Local Similarity	50.0%;	Pred. No. 99;		
Matches 11;	Conservative 2;	Mismatches 7;	Indels 2;	Gaps 1;
3 HHFRGT--VHVGTKITRLVLTG 22				

Db	132	HEALMGLQHLHVNKT1HRDVK	153
RESULT	14		
Q9UB86			
ID	Q9UB86	PRELIMINARY;	PRT; 202 AA.
AC	Q9UB86;		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DT	01-OCT-2001	(TREMBLrel. 18, Last annotation update)	
DE	C-Reactive protein (Fragment).		
OS	Tachypleus tridentatus (Japanese horseshoe crab).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;		
OC	Limulidae; Tachypleus.		
OX	NCBI_TaxID=6853;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=HEPATOPANCREAS;		
RA	MEDLINE=99421634; PubMed=10491075;		
RA	Iwaki D., Osaki T., Mizunoe Y., Wai S. N., Iwanaga S., Kawabata S.;		
RT	"functional and structural diversities of C-reactive proteins present		
RT	in horseshoe crab hemolymph plasma."		
LT	Eur. J. Biochem. 264:314-326(1999).		

DR	InterPro: IPR001759; Pentaxin.
DR	Pfam: PF000354; pentaxin.1
DR	PRINTS: PR00895; PENTAXIN.
DR	ProDom: PD002153; Pentaxin.1
DR	SMART: SM00159; PTX.1
DR	PROSITE: PS00289; PENTAXIN.1
FT	NON_TER 1
FT	NON_TER 202
SQ	SEQUENCE 202 AA; 22780 MW; 8C5CC315BF2FB126 CRC64;
Query Match 38.8%; Score 47; DB 5; length 202;	
Best Local Similarity 47.4%; Pred. No. 14;	
Matches	9; Conservative 3; Mismatches 7; Indels 0; Gaps 0
DY	4 HIRPGIVHGKTIHRLVTS 22
db	182 IIGVNIHHGKTFQFRVYDG 200

```

RESULT 15
08S810
ID 08S810 PRELIMINARY; PRT; 476 AA.
AC 08S810;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Putative prolylcarboxypeptidase.
GN A7G524280.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxId=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RL Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
[2] Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Town C.D., Kaul S.;

```

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005967; AAM15036.1; -.
KM Carboxypeptidase.
SQ SEQUENCE 476 AA; 54423 MW; 3C3EF8E1F8A580D2 CRC64;

Query Match 38.8%; Score 47; DB 10; Length 476;
Best Local Similarity 42.9%; Pred. No. 34;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 FFHIFRGIVHVGKTIHRLVT 21
: || || | : : : ||
Db 410 YFHQIFRVLKNISSIVALVT 430

Search completed: May 23, 2003, 08:56:43
Job time : 27.5 secs